DOCUMENT RESUME

ED 094 890 PS 007 447

AUTHOR	Cline, Marvin G.; And Others
TITLE	Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volume 1A: Early Effects of Follow Through. Final Report.
INSTITUTION	Abt Associates, Inc. Cambridge, Mass.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
REPORT NO	AAI-74-13
PUB DATE	1 Mar 74
CONTRACT	OEC-0-72-5221
NOTE	380p.; For Volume 1B, see PS 007 448
EDRS PRICE	MF-$0.75 HC-$18.60 PLUS POSTAGE
DESCRIPTORS	*Academic Achievement; Administrative Problems; Attitudes; Compensatory Education Programs; Intervention; *Models; Parental Background; *Primary Education; *Program Evaluation; Racial Differences; Sex Differences; Social Differences; Standardized Tests; *Statistical Analysis; Teacher Characteristics
IDENTIFIERS	*Project Follow Through

ABSTRACT

The design of this Planned Variation study examines the impact of Project Follow Through on children, focusing on the 1-year kindergarten experiences of the third Follow Through group. Chapter 1 presents a short history of Follow Through and a description of each of the participating program sponsors. Chapter 2 considers the problems faced in constructing a manageable set of questions which could be put to the available data. Chapter 3 describes the subject of sites and children utilized in the analyses. Chapter 4 describes the covariables used in making the adjustments for initial differences between groups being compared. Chapter 5 presents the statistical strategy chosen for these analyses, the methods of presenting results, and the manner of interpreting the tabulated results. Chapter 6 represents a pause in the flow of the evaluation report to provide the reader with some contextual information necessary for sensing the meaning behind some of the numbers reported. Chapter 7 presents the major comparisons between the Project Follow Through and the non-Follow Through schools across all programs, and by each program. Chapter 8 presents a series of studies which suggests some interesting educational implications, while chapter 9 considers the problem of comparing the several programs on the outcome measures. (CS)

ED 094890

ABT ASSOCIATES INC.
55 WHEELER STREET, CAMBRIDGE, MASSACHUSETTS 02138
TELEPHONE • AREA 617-492-7100
TELEX: 710-320-6367

Prepared for:

THE U.S. OFFICE OF EDUCATION
Office of Planning, Budgeting, and Evaluation
Department of Health, Education, and Welfare

Contract No. OEC-0-72-5221

PS 007442

FINAL REPORT

EDUCATION AS EXPERIMENTATION: EVALUATION OF THE FOLLOW THROUGH PLANNED VARIATION MODEL

VOLUME IA: Early Effects of Follow Through

Marvin G. Cline
Project Director

Ames, N., Anderson, R., Bale, R., Ferb, T., Joshi, M., Kane, M., Larson, J., Park, D., Proper, E., Stebbins, L., Stern, C.

Report No. AAI-74-13 | March 1, 1974

EDUCATION AS EXPERIMENTATION:

EVALUATION OF THE FOLLOW THROUGH PLANNED VARIATION MODEL

VOLUME: 1A

Early Effects of Follow Through

VOLUME 1A

	Page
I. AN OVERVIEW: THE GROWTH OF THE PLANNED VARIATION MODEL	I-1
1.0 INTRODUCTION	I-1
2.0 ORIGINS	I-1
3.0 PLANNED VARIATION EXPERIMENTATION THROUGH PROGRAM SPONSORS	I-4
4.0 PROGRAM IMPLEMENTATION	I-5
5.0 FOLLOW THROUGH PROGRAM SPONSORS	I-7
5.1 Sponsor 2: Far West Laboratory	I-8
5.2 Sponsor 3: University of Arizona	I-9
5.3 Sponsor 5: Bank Street College	I-10
5.4 Sponsor 7: University of Oregon	I-11
5.5 Sponsor 8: University of Kansas	I-12
5.6 Sponsor 9: High/Scope Educational Research Foundation	I-13
5.7 Sponsor 10: University of Florida	I-14
5.8 Sponsor 11: Educational Development Center	I-15
5.9 Sponsor 12: University of Pittsburgh	I-16
5.10 Sponsor 14: Southwest Educational Development Laboratory	I-17
II. GOALS OF THE ANALYSIS	II-1
1.0 INTRODUCTION	II-1
2.0 LIMITATIONS AND QUALIFICATIONS	II-2
2.1 Prematurity	II-2
2.2 Limitations of the Follow Through Design	II-5
2.2.1 Imbalance	II-5
2.2.2 Purposive Selection	II-7
2.2.3 Sponsor Diversity	II-9
2.3 Practical Limitations	II-11
2.3.1 Implementation	II-11
2.3.2 Methodological Limitations	II-13
3.0 THE ANALYSIS STRATEGY	II-17
3.1 FT/NFT Contrasts	II-17
3.2 Big City Effect on FT/NFT Contrasts	II-18
3.3 Influence of Relevant Variables on FT/NFT Contrasts	II-19
3.4 Influence of Curriculum Models on FT/NFT Contrasts	II-22
3.5 Description of Follow Through Teachers	II-23
3.6 Description of Follow Through Parents	II-23
3.7 Description of Selected Sponsors' Implementation Problems	II-24

III. DESCRIPTION OF THE ANALYTIC SUBSET III-1

1.0 INTRODUCTION III-1

Sponsor Identification Code III-1a

2.0 CHARACTERISTICS III-1

2.1 Geographic Distribution III-3

2.1.1 Sponsor 2: Far West Laboratory III-7
2.1.2 Sponsor 3: University of Arizona III-8
2.1.3 Sponsor 5: Bank Street College III-8
2.1.4 Sponsor 7: University of Oregon III-8
2.1.5 Sponsor 8: University of Kansas III-9
2.1.6 Sponsor 9: High/Scope Foundation III-9
2.1.7 Sponsor 10: University of Florida III-10
2.1.8 Sponsor 11: Educational Development Center III-10
2.1.9 Sponsor 12: University of Pittsburgh III-10
2.1.10 Sponsor 14: Southwest Educational Development Laboratory III-11

2.2 Demographic Descriptions III-11

3.0 OUTCOME VARIABLES III-14

3.1 Academic Achievement III-14
3.2 Motivational Orientation III-15
3.3 Absence III-16

IV. COVARIATES IV-1

1.0 INTRODUCTION IV-1

2.0 WHICH COVARIATES? IV-2

2.1 Levels of Analysis and the Meaning of Covariates IV-4

3.0 THE COVARIATES USED IN OUR ANALYSIS IV-5

V. METHODOLOGICAL PROBLEMS IN THE ANALYSIS OF THE FOLLOW THROUGH DATA V-1

1.0 THE CAVEATS V-1

1.1 The Covariance Technique V-1
1.2 Imbalance V-2
1.3 Missing and Incomplete Data V-2
1.4 Fallibility of Measures V-3
1.5 Inhomogeneity V-3
1.6 Non-Probabilistic Sampling V-3

2.0 METHODS SELECTED FOR THE CURRENT ANALYSIS V-4

2.1 Effects of Aggregation V-5

3.0 PRESENTATION OF RESULTS V-7

	Page
4.0 NOMINAL CODING SCHEMES	V-8
4.1 Classical Approach	V-9
4.2 Rationale for Our Scheme	V-10
4.3 An Example	V-10
4.4 Generalizations	V-11
5.0 THE CHOICE OF EFFECTS	V-12
VI. AN OVERVIEW OF PARALLEL STUDIES	VI-1
1.0 SUMMARY: PARENT STUDIES	VI-1
2.0 SUMMARY: TEACHER STUDIES	VI-3
3.0 SUMMARY: IMPLEMENTATION STUDY	VI-5
VII. SUMMARY: FT/NFT CONTRASTS	VII-1
1.0 THE ONE YEAR KINDERGARTEN STUDY	VII-2
1.1.0 INTRODUCTION	VII-2
1.2.0 METHOD	VII-2
1.2.1 Design	VII-2
1.2.2 Analytic Subset	VII-3
1.2.3 The Variables	VII-3
1.2.4 The Power of the Analyses	VII-9
1.3.0 PATTERNS OF EFFECTS	VII-9
1.3.1 Main Effects	VII-9
1.3.2 Sponsor Effects	VII-11
1.3.3 Effects of Covariance Adjustment	VII-23
1.4.0 THE EFFECT OF PRETEST DELAY	VII-27
1.5.0 DISCUSSION	VII-30
2.0 BIG CITY STUDY	VII-31
2.1.0 INTRODUCTION	VII-31
2.2.0 METHOD	VII-31
2.2.1 Analytic Subset	VII-31
2.2.2 Analytic Design	VII-34
2.2.3 Variables	VII-34
2.3.0 RESULTS	VII-34
2.3.1 Main Effects	VII-34
2.3.2 Sponsor Effects	VII-36
2.4.0 DISCUSSION	VII-36
3.0 RELATED DATA	VII-45
3.1.0 SOCIOECONOMIC STATUS AND PUPIL ACHIEVEMENT	VII-45
3.2.0 TIME OF TESTING CORRELATIONS	VII-48

	Page
4.0 SPONSOR VIGNETTES	VII-53
4.0.1 INTRODUCTION	VII-53
4.0.2 LEVELS OF ANALYSIS	VII-55
4.0.2.1 School Level	VII-55
4.0.2.2 Class Level	VII-57
4.0.2.3 Child Level	VII-57
4.0.3 METHODS	VII-58
4.1.0 SPONSOR 2: FAR WEST LABORATORY	VII-59
4.2.0 SPONSOR 3: UNIVERSITY OF ARIZONA	VII-66
4.3.0 SPONSOR 5: BANK STREET COLLEGE	VII-73
4.4.0 SPONSOR 7: UNIVERSITY OF OREGON	VII-81
4.5.0 SPONSOR 8: UNIVERSITY OF KANSAS	VII-88
4.6.0 SPONSOR 9: HIGH/SCOPE EDUCATIONAL RESEARCH FOUNDATION	VII-95
4.7.0 SPONSOR 10: UNIVERSITY OF FLORIDA	VII-102
4.8.0 SPONSOR 11: EDUCATIONAL DEVELOPMENT CENTER	VII-109
4.9.0 SPONSOR 12: UNIVERSITY OF PITTSBURGH	VII-116
4.10.0 SPONSOR 14: SOUTHWEST EDUCATIONAL DEVELOPMENT LABORATORY	VII-124
5.0 COHORT I AND MULTIPLE COHORT STUDIES	VII-133
5.1.0 INTRODUCTION, COHORT I STUDY	VII-133
5.2.0 METHOD	VII-133
5.2.1 Analytic Subset	VII-133
5.2.2 Design	VII-134
5.3.0 RESULTS	VII-138
5.4.0 DISCUSSION	VII-140
5.5.0 INTRODUCTION, MULTIPLE COHORT STUDY	VII-144
5.6.0 METHOD	VII-144
5.6.1 Analytic Subset	VII-144
5.6.2 Design	VII-144
5.7.0 RESULTS	VII-150
5.8.0 DISCUSSION	VII-150
6.0 CONCLUDING STATEMENTS	VII-154
VIII. CLASS AND CHILD VARIABLES AFFECTING FT/NFT CONTRASTS	VIII-1
1.0 CLASS ETHNIC COMPOSITION	VIII-1
1.1.0 INTRODUCTION	VIII-1

	Page
1.2.0 METHOD	VIII-2
1.2.1 Analytic Subset	VIII-2
1.2.2 Measures	VIII-3
1.2.3 Analytic Method	VIII-3
1.3.0 RESULTS	VIII-6
1.4.0 DISCUSSION	VIII-11
2.0 ENTRY LEVEL STUDY	VIII-13
2.1.0 INTRODUCTION	VIII-13
2.2.0 METHOD	VIII-14
2.2.1 Analytic Subset	VIII-14
2.2.2 Measures	VIII-14
2.2.3 Analytic Method	VIII-14
2.3.0 RESULTS	VIII-16
2.4.0 DISCUSSION	VIII-27
3.0 CHILD LEVEL PRESCHOOL STUDY	VIII-29
3.1.0 INTRODUCTION	VIII-29
3.2.0 METHOD	VIII-32
3.2.1 Subjects	VIII-32
3.2.2 Measures	VIII-32
3.2.3 Analytic Plan	VIII-34
3.3.0 RESULTS	VIII-35
3.4.0 DISCUSSION	VIII-40
4.0 CHILD ETHNICITY STUDY	VIII-42
4.1.0 INTRODUCTION	VIII-42
4.2.0 METHOD	VIII-44
4.2.1 Subjects	VIII-44
4.2.2 Measures	VIII-44
4.2.3 Analytic Plan	VIII-45
4.3.0 RESULTS	VIII-46
4.4.0 DISCUSSION	VIII-52
5.0 CHILD STUDY OF SEX DIFFERENCES	VIII-54
5.1.0 INTRODUCTION	VIII-54
5.2.0 METHOD	VIII-56
5.2.1 Subjects	VIII-56
5.2.2 Measures	VIII-56
5.2.3 Analytic Plan	VIII-58
5.3.0 RESULTS	VIII-59
5.4.0 DISCUSSION	VIII-60
6.0 CONCLUDING STATEMENTS	VIII-61

		Page
IX.	CROSS SPONSOR COMPARISONS: CONCLUDING REMARKS	IX-1
	1.0 INTRODUCTION	IX-1
	2.0 ETHNICITY	IX-2
	3.0 HEAD START	IX-8
	4.0 ENTRY LEVEL	IX-11
	5.0 GUMPGOOKIES	IX-13
	6.0 TIME OF TESTING	IX-18
	7.0 PLANS FOR THE FUTURE	IX-23
X.	SUMMARY (This summary will be bound separately)	
	BIBLIOGRAPHY	

List of Tables

Volume IA - Early Effects of Follow Through

Table		Page
III-1	The Number and Proportion of Schools/Classes/Children for Each Sponsor Within Each Region by FT/NFT	III-5
III-2	The Number and Proportion of Schools/Classes/Children for Each Sponsor Within Each City Size by FT/NFT. . . .	III-6
III-3	School/Class/Child Sample Size and Means on Four Covariates for Each Sponsor by FT/NFT	III-12
IV-1	Overview of Variables Used as Covariates in Studies at Three Levels of Analysis	IV-6
VII-1	Distribution of the FT and NFT Schools Across Ten Sponsors for the Study of One Year Kindergarten Effects	VII-4
VII-2	Predictor Coding Scheme for the Nested Analysis of One Year Kindergarten Effects	VII-6
VII-3	Contrast Coding Scheme for the Factorial Analysis of One Year Kindergarten Effects.	VII-7
VII-4	Effects of Covariance Adjustment in Criterion Standard Deviation Units.	VII-24
VII-5	Zero-Order Correlations of Pretest Delay with Fall WRAT Scores by Sponsor by FT/NFT.	VII-28
VII-6	Distribution of the Complete Analytic Subset of FT and NFT Schools Across Ten Sponsors.	VII-32
VII-7	Distribution of Big City Schools Across Sponsor's FT and NFT Populations.	VII-33
VII-8	Zero-Order Correlations Between SES Indicators and Fall and Spring WRAT.	VII-46
VII-9	Partial Correlations of Spring Scores with Length of Instructional Interval Controlling for Fall WRAT, FT Schools.	VII-49, IX-19
VII-10	Partial Correlations of Spring Scores with Length of Instructional Interval Controlling for Fall WRAT, NFT Schools	VII-50, IX-20

Table		Page
VII-11	Distribution of Schools in the Three Year Longitudinal Effects Study Population by Sponsor by FT/NFT	VII-135
VII-12	Contrast Coding Scheme for the Factorial Analysis of Three Year Effects Study	VII-136
VII-13	Contrast Coding Scheme for the Nested Analysis for Three Year Effects Study.	VII-137
VII-14	Partition of Variance for the Three Year Effects Study	VII-139
VII-15	Distribution of Schools in the Multiple-Cohort Study Population by Sponsor and by Follow Through Participation Status.	VII-145
VII-16	Contrast Coding Scheme for the Sponsor by FT/NFT Interaction Analysis (A) of the Multiple-Cohort Study	VII-147
VII-17	Contrast Coding Scheme for the Sponsor by FT/NFT by Time Analysis (B) of the Multiple-Cohort Study . . .	VII-148
VII-18	Statistics from the Regression Analyses of Spring WRAT Scores for the Multiple-Cohort Study	VII-149
VII-19	Statistics from Regression Analysis (B) for the Multiple-Cohort Study with Pretest Delay Included in Set G.	VII-153
VIII-1	F Ratios and Factors in the Hierarchical Model of ANCOVA Ethnic Composition Study.	VIII-5
VIII-2	Number and Percent of Total Numbers of Classes Falling Into Each of Five Categories of Precent White in Class for Each Sponsor's FT and NFT Group and for the Pooled FT and NFT Groups.	VIII-7
VIII-3	Summary of Effects, Class Level Analysis.	VIII-8
VIII-4	F Ratios and Factors in the Hierarchical Model of ANOVA and ANCOVA - Entry Level Study.	VIII-15
VIII-5	Class Entry - Level Study: Variance Accounted for by the Interaction of Initial Achievement Level, Sponsor Membership, FT/NFT Membership; Total Variance Accounted for, F Ratio, and Significance for the ANCOVA Models	VIII-17

List of Tables - Continued

Table		Page
VIII-6	The Distribution of Sponsor's FT/NFT Classes Across Five Categories of Initial Achievement Defined in Terms of the Sample Mean (35.68) and Standard Deviation (6.19) of Fall WRAT.	VIII-18
VIII-7	Fall WRAT Means, Standard Deviations and Number of Classes for Each Sponsor's Follow Through/non-Follow Through Groups.	VIII-19
VIII-8	Distribution of Subjects for Head Start by Sponsor by FT/NFT Analysis with Entering WRAT, Poverty Level, and Mother's Education Group Mean	VIII-33
VIII-9	Distribution of Subjects for Ethnic Type by Sponsor by FT/NFT Analysis with Entering WRAT, Poverty Level, and Mother's Education Group Mean.	VIII-43
VIII-10	Distribution of Subjects for Sex by Sponsor by FT/NFT Analysis with Entering WRAT, Poverty Level, and Mother's Education Group Mean.	VIII-57
IX-1	Sponsor Effects on Four Outcomes Ethnicity Study	IX-4
IX-2	Sponsor Effects on Three Outcomes Head Start Study . . .	IX-4
IX-3	Sponsor Effects on Two Outcomes Entry Level Study. . . .	IX-12

List of Figures

Volume IA - Early Effects of Follow Through

Figures

		Page
III-1	Number of Sites for Each Sponsor Within Each Region and City Size.	III-4
VII-1	Follow Through Effects Profile for Main Effect (10 Sponsor School Population Excluding Big City) . . .	VII-10
	Excluding Big City applies to Figures VII-2 through VII-11	
VII-2	Follow Through Effects Profile for Sponsor 2.	VII-12,61
VII-3	Follow Through Effects Profile for Sponsor 3.	VII-13,68
VII-4	Follow Through Effects Profile for Sponsor 5.	VII-14
VII-5	Follow Through Effects Profile for Sponsor 7.	VII-15
VII-6	Follow Through Effects Profile for Sponsor 8.	VII-16
VII-7	Follow Through Effects Profile for Sponsor 9.	VII-17
VII-8	Follow Through Effects Profile for Sponsor 10	VII-18
VII-9	Follow Through Effects Profile for Sponsor 11	VII-19
VII-10	Follow Through Effects Profile for Sponsor 12	VII-20,118
VII-11	Follow Through Effects Profile for Sponsor 14	VII-21
VII-12	Effects of Covariance on the WRAT Follow Through Effects of Ten Sponsors and Their Aggregate	VII-26
VII-13	Follow Through Effects Profile for Main Effects (10 Sponsor School Population Including Big Cities) . .	VII-35
	Including Big City applies to Figures VII-14 through VII-20	
VII-14	Follow Through Effects Profile for Sponsor 5.	VII-38,74
VII-15	Follow Through Effects Profile for Sponsor 7.	VII-39,82
VII-16	Follow Through Effects Profile for Sponsor 8.	VII-40,90
VII-17	Follow Through Effects Profile for Sponsor 9.	VII-41,96
VII-18	Follow Through Effects Profile for Sponsor 10	VII-42,103
VII-19	Follow Through Effects Profile for Sponsor 11	VII-43,110

Figure		Page
VII-20	Follow Through Effects Profile for Sponsor 14.	VII-44,126
VII-21	Class Study Effects Profile for Sponsor 2, Cohort III, Kindergarten	VII-62
VII-22	Child Study Effects Profile for Sponsor 2, Cohort III, Kindergarten	VII-64
VII-23	Class Study Effects Profile for Sponsor 2, Cohort III, Kindergarten	VII-69
VII-24	Child Study Effects Profile for Sponsor 3, Cohort III, Kindergarten	VII-71
VII-25	Class Study Effects Profile for Sponsor 5, Cohort III, Kindergarten	VII-76
VII-26	Child Study Effects Profile for Sponsor 5, Cohort III, Kindergarten	VII-78
VII-27	Class Study Effects Profile for Sponsor 7, Cohort III, Kindergarten	VII-84
VII-28	Child Study Effects Profile for Sponsor 7, Cohort III, Kindergarten	VII-86
VII-29	Class Study Effects Profile for Sponsor 8, Cohort III, Kindergarten	VII-91
VII-30	Child Study Effects Profile for Sponsor 8, Cohort III, Kindergarten	VII-93
VII-31	Class Study Effects Profile for Sponsor 9, Cohort III, Kindergarten	VII-98
VII-32	Child Study Effects Profile for Sponsor 9, Cohort III, Kindergarten	VII-99
VII-33	Class Study Effects Profile for Sponsor 10, Cohort III, Kindergarten	VII-105
VII-34	Child Study Effects Profile for Sponsor 10, Cohort III, Kindergarten	VII-107
VII-35	Class Study Effects Profile for Sponsor 11, Cohort III, Kindergarten	VII-112
VII-36	Child Study Effects Profile for Sponsor 11, Cohort III, Kindergarten	VII-114

List of Figures - Continued

Figure		Page
VII-37	Class Study Effects Profile for Sponsor 12, Cohort III, Kindergarten	VII-120
VII-38	Child Study Effects Profile for Sponsor 12, Cohort III, Kindergarten	VII-121
VII-39	Class Study Effects Profile for Sponsor 14, Cohort III, Kindergarten	VII-128
VII-40	Child Study Effects Profile for Sponsor 14, Cohort III, Kindergarten	VII-130
VII-41	Effects Profile for Three Year Longitudinal Study: MAT Spelling	VII-141
VII-42	Effects Profile for Three Year Longitudinal Study: MAT Reading	VII-142
VII-43	Effects Profile for Three Year Longitudinal Study: MAT Arithmetic	VII-143
VII-44	Comparison of Cohort I and Cohort III One Year Kindergarten Gains on WRAT Scores for Follow Through and non-Follow Through Schools in Six Sponsors	VII-151
VIII-1a, 1b, 1c	Main Effects of Classroom Ethnic Composition Across All Sponsors for Locus Positive, Gumpgookies and Absence	VIII-9
VIII-2	Regressions of Spring WRAT on Entry Level for Each Sponsor, by Follow Through and non-Follow Through . . .	VIII-21
VIII-3	Regressions of MAT Reading on Entry Level for Each Sponsor, by Follow Through and non-Follow Through . . .	VIII-22
VIII-4	Adjusted Spring WRAT Scores for Cohort III, Kindergarten	VIII-37
VIII-5	Adjusted Spring MAT Math Scores for Cohort III, Kindergarten	VIII-38
VIII-6	Preschool by Sponsor by FT/NFT (MAT Reading)	VIII-39
VIII-7	Ethnicity by Sponsor by Treatment Interactions for 10 Sponsors on Adjusted Spring WRAT Scores	VIII-47
VIII-8	Ethnicity by Sponsor by Treatment Interactions for 10 Sponsors on Adjusted Spring Math Scores	VIII-48

List of Figures - Continued

Figure		Page
VIII-9	Ethnicity by Sponsor by Treatment Interactions for 10 Sponsors on Adjusted Spring Listening for Sounds Scores	VIII-49
VIII-10	Ethnicity by Sponsor by Treatment Interactions for 10 Sponsors on Adjusted Spring PPVT Scores	VIII-50
IX-1	Follow Through Effects Profile for Gumpgookies-Excluding Big Cities	IX-15
IX-2	Follow Through Effects Profile for Gumpgookies-Including Big Cities	IX-16

PREFACE

This is the first in the Abt Associates Inc. series of reports on the impact of Project Follow Through. As the experimental phase of this multifaceted attempt to change the character of the elementary grades in the American school system draws to a close, the evaluative phase of the program becomes more intense. Each year, for the past four years, a large group of children from almost every major segment of the United States started their public school careers in classrooms which were under the supervision of one of the several constituent programs of Follow Through. The expectation is that a large number of these children will continue through the third grade in classrooms supervised by the same educational program. As each group of children graduates the third grade, a full set of data covering the whole span of involvement with Follow Through becomes available for analysis. At the time the data for this report were received for analysis, the first groups of children to enter kindergarten were in the midst of their third grade year, and the third group of entering kindergartners had just completed their kindergarten year. Thus, the data analyzed for this first Abt report do not include any children who completed the full four-year course of Follow Through. In fact, for a variety of reasons, the major emphases of this report are on the one-year kindergarten experiences of the third group. However, the third grade data for the first group are now at hand, and will be analyzed for the report to be submitted a year from now. The second Abt report will continue to focus on the third group of children to enter the program, whose first grade data are also now at hand. The fourth and last group entered kindergarten as part of the Follow Through experiment in September 1973, and have at this writing three and a half years to go before completing the full Follow Through course. By the time this last group graduates from Follow Through, three successive sets of kindergarten through third grade data will have been analyzed and compared to each other. Strictly speaking, these sets of data are not directly comparable since each represents a very different group of children, and programs operating under very different conditions. Nevertheless, the goal is to compare the four year

patterns to each other so that the stability of the pattern for each educational program can be assessed. Clearly this means that as the operational side of the Follow Through programs phases out, the magnitude of relevant data increases along with the complexity of the analyses. This first report, therefore, involves not only the smallest subset of data to be included in any of the reports, but it also involves the least complex set of analyses.

This also means that the results reported in this first volume are necessarily tentative. This is not because the analyses are tentative, but because the definitive analyses are not yet possible. This poses a conflict for the evaluators whose responsibility includes the provision of results which are capable of contributing to policy decisions. As social scientists, it is obvious to us that this first report based upon an analysis of kindergarten data cannot support policy decisions about a four year program. As responsible evaluators, we also know that decisions must be made quickly, often before analyses can be completed. The temptation is to look into the results, even the very earliest results, for hints or leads which might give a sense of where the data are going. Unfortunately this kind of situation is extremely dangerous because it leads to a proliferation of self-fulfilling prophecies. The prophecy that some approaches to elementary education are having minimal effects may lead some policymakers to the premature fulfillment of that prophecy through the elimination of those approaches from the experiment. Drastic steps such as total elimination from the experiment are not, however, necessary to move such a prophecy along to fulfillment. To be singled out, no matter how unfairly or prematurely, for failing to produce large changes in the academic achievement history or in the motivational status of kindergartners is debilitating to the morale of the staff and destructive to the relationship between the programs and the communities with whom the programs work. These are factors which can have destructive influences upon the future development of the program and the children involved, thereby speeding the prophecy to fulfillment. For such reasons, we have chosen to remain on the conservative side of the conflict between our social science and evaluation obligations. We have resisted drawing many conclusions about the value or lack of value of many aspects of the Follow Through programs. The goal of this first report

is not, therefore, to be heavily interpretive. It is rather to describe the nature and extent of the Follow Through effects as they are revealed in our analyses of the first set of kindergarten data. There are such effects; Follow Through is having an impact on the world of elementary education, but it is a complex effect varying according to the kind of program, the kinds of children, and the conditions and places of application. This should not be surprising to those who devised the Planned Variation experiment, and it should certainly not be surprising that the evaluators of the program treat these findings as early effects in need of expansion and replication before being submitted as evidence for decision making. We are reporting here that there is indeed a true Follow Through effect present in the kindergarten test scores, and that some few of the conditions under which the effect emerges are beginning to be identified. But it is not yet time to start drawing conclusions about educational practices at this stage of the study.

There is another problem which we as evaluators faced and will continue to face throughout the years of this study. The problem stems from the fact that we entered the Follow Through evaluation late in the history of the program. Our contract to analyze the longitudinal data base started in July 1972. The first set of data was received from the Stanford Research Institute (the agency responsible for, among other tasks, the collection and encoding of data relevant to the national longitudinal study) in October 1972. These data were the basis of an interim report submitted to the Office of Education on January 31, 1973. On January 15, 1973, the first full set of data was received. The analyses of these data, which were started six weeks after receipt and completed during the summer of 1973, produced the findings contained in this report. The purpose of this chronology is to make clear the problem we as evaluators face and which appears to be the bane of most evaluation efforts. By the time we assumed our evaluative responsibilities, the design of the program had been set and in operation; the data had been collected for years without regard for a well established analytic plan; some events which might have been useful in interpreting the findings were not recorded and are long since forgotten by the individuals concerned; and the battery of

instruments selected in the past now presents restrictions to the kinds of questions which we might like to ask. We entered the study when it was well under way, and although there is some flexibility in our reporting schedule, our contractual obligations hardly allow us the luxury of contemplation which a task of this nature requires. This appears to be the typical situation faced by program evaluators which reflects the general status of evaluation in education today. Despite this unfortunate situation, we assume full responsibility for the many inadequacies in the present Report. We only wish that fewer of these inadequacies resulted from our after-the-fact relationship to the programs and the evaluation design.

Given our overt decision to avoid premature interpretations and not to attempt to tie all data together for a full picture of each experimental model for this Report, it is entirely possible that covert interpretations will emerge from such a complex set of data and findings. Research is not a value-free process, and evaluation (which is so heavily tied to decision making) is all the more enmeshed in political processes. There is every reason to expect that the biases of the evaluators will be found throughout a report which is designed to be incomplete in its conclusions. The awareness of the evaluators of this tendency is one way to prevent massive distortions, and we believe that we are aware of ours. A full reporting of all relevant information so that the reader may judge the extent to which biases are operating is a way to rectify some of the distortions, and we have attempted, to the point of being deliberately redundant in the presentation of information, to report fully. There is one further preventative step to take and that is to assert our biases in advance and let the reader beware. We would like to see the Follow Through programs work. We hope that the many approaches to the reformation of elementary education will significantly alter the educational history of the participating children because we believe that the non-Follow Through world of elementary education is in part responsible for the relatively poor performance of many of the children of poverty. We believe that the Follow Through programs are not only introducing new styles of instruction to the public schools, but they are also introducing

new goals for the educational process. They are changing the traditional decision making processes in schools, and they are ligitimizing more contemporary notions of child development in the eyes of schoolmen. They are, in other words, true agents of change whose status external to the public schools needs to be nurtured since such far reaching changes are not likely to be maintained without their constant prodding.

These are strong, and perhaps overstated biases, and they must be laid out so that the reader will be sensitive to their potential. We have tried to suspend them in preparing this Report, but it is for the reader to judge the extent to which we have failed.

The plan of this Report needs to be stated here as an aid in dealing with such a massive set of data.

Chapter I presents a short history of Follow Through and a description of each of the participating program Sponsors. A summary of the Sponsor descriptions will be repeated later when a summary of some findings for each Sponsor is presented so that the reader will not have to go back to Chapter I to recall the relevant background information when the findings are considered.

Chapter II considers the problems faced in constructing a manageable set of questions which could be put to these data. The overall analytic strategy consists of the major questions which were selected for examination and these questions are stated in this chapter.

Chapter III describes the subset of sites and children utilized in the analyses. We have deliberately chosen to refer to the groups included in the analyses as subsets of the total Follow Through population rather than to use the term sample, because sampling criteria were judgmental, they varied from site to site, and they were not designed to be representative of Follow Through. It is critical, therefore, that a description of the subsets which were included in the analyses be presented here, and be kept in mind by the reader whenever findings are presented. In order to facilitate this, we chose to repeat relevant sections of the subset descriptions when some of the findings were summarized.

Chapter III also includes a description of the battery of instruments used in the study.

Chapter IV describes the covariables used in making the adjustments for initial differences between groups being compared. These covariables constitute rival hypotheses, and the reader must be clearly aware of which hypotheses we have attempted to rule out and which were not dealt with.

Chapter V presents the statistical strategy chosen for these analyses, the methods of presenting results, and the manner of interpreting the tabulated results. For a more complete description of the general linear model, which is the basis of our statistical strategy, the reader is referred Volume IB of this Report.

Chapter VI represents a pause in the flow of the evaluation report to provide the reader with some contextual information necessary for sensing the meaning behind some of the numbers reported. Here we have summarized three small studies on teachers, parents, and the problems of implementing the models faced by the program Sponsors. These studies are reported in Volume IB as separate monographs because they have not yet been merged with pupil data. But they provide a good deal of information on the extra-classroom events faced by each of the programs and represent, therefore, the context for the educational activities which constitute the Follow Through experiment. Before the findings are presented, we considered it essential that the reader have some feeling for these factors, but we did not want to require that the full studies be read before coming to the findings. Thus, we have interjected a short summary of these studies in Chapter VI, and refer the reader to Volume IB for the full reports.

Chapter VII presents the major comparisons between the Follow Through and the non-Follow Through schools across all programs, and by each program. Several small studies bearing on the question of a Follow Through effect on kindergarten children, and a summary of findings on some of the earlier groups of children to have gone through the programs are also presented here. In order to provide an initial picture of the pattern of effects for each program, a series of program vignettes is presented in this chapter which brings together a summary of the goals of the program, some

properties of the subset of the sites and children involved in the analyses, and the more important findings for that program. We have emphasized throughout this Report that it is premature to draw definitive conclusions about program impacts on children, so we have simply summarized these data for each program in vignette form. We shall begin our interpretative tasks in the next annual Report when these patterns can be considered for their longitudinal stability and therefore can justifiably be interpreted for their educational significance.

Chapter VIII presents a series of studies which we expect will lead to the most important of the educational implications of these programs. These studies examine some of the conditions under which the several program effects were obtained. Here we have examined a number of types of classes and properties of children as these interact with the Follow Through programs to produce effects in achievement and motivational measures.

Chapter IX considers the problem of comparing the several programs on the outcome measures. The issue here is to estimate the extent to which educational conclusions can be drawn at this point in the longitudinal study. The plans for the next set of analyses are also presented in this chapter.

The Summary, which is designed to highlight selected aspects of the findings, is bound separately.

Acknowlegments

The very large number of components, necessarily contained in a study of Follow Through, requires that a large number of individuals make major contributions. At the same time, the complexity of the work means that it is often difficult to separate out the unique contribution of each individual. This Report was produced jointly by a group of individuals who interacted a great deal about each others' work. We, the Abt Associate's Follow Through staff, are collectively responsible for the Report, but we have also attempted to put our individual stamps on the total package. As a group therefore, we would like to acknowledge here the very great support given us by a variety of advisors in and out of the U.S. Office of Education.

First, we would like to thank Dr. Garry McDaniels, the former Chief of Evaluation, Follow Through Branch, who served as our Project Monitor until April, 1973. Dr. McDaniels saw us through the first months of settling into the data and translating our original plans into operational procedures. His sensitivity to the nature of the problem and the needs of social scientists as they bring the problem into focus was critical to our work and is deeply appreciated. Mr. Eugene Tucker serves as Dr. McDaniels successor and he remains in that position now that the Office of Follow Through evaluation has been transferred to the Office of Planning, Budgeting and Evaluation of the U.S.O.E. Mr. Tucker has been extremely helpful during the second half of our task: reporting the results of the first study. His critical comments have been incisive and useful and his willingness and ability to translate the concerns of the U.S.O.E. into social science terms has made our task considerably more manageable.

Next, a group of consultants has provided major input to our thinking and has served as a sounding board for our plans. Foremost in the area of methodological support is Dr. Jacob Cohen of New York University who has spent long hours critically examining our procedures.

Dr. William Meyer of Syracuse University has served as our External Reviewer and has examined in detail all of the products associated with our work. Both of these consultants have had major impact on our thinking, but, we of course remain responsible for the Report.

The remaining members of our panel of consultants are: Dr. E. Kuno Beller, Temple University; Dr. David C. Berliner, Far West Regional Laboratory; Dr. Robert Brennan, SUNY at Stony Brook; Dr. Donald Campbell, Northwestern University; Dr. Jimmie C. Fortune, Virginia Polytechnic Institute; Dr. Edward J. O'Connell, Syracuse University. The panel members have met with us regularly both as a group and individually and they have provided severe criticisms, useful suggestions, and the encouragement that is so important, but which only truly respected colleagues can supply.

It is also appropriate at this point to acknowledge the contribution made by the members of the staff and the consultants of the Huron Institute which serves as technical advisors to the Office of Evaluation of Follow Through. Dr. Marshall Smith, Harvard University; Dr. Andrew Porter, Michigan State University; Jane David and Richard Elmore both of Huron Institute, participated in one or more of the several planning sessions as advisors to Follow Through. They have also read various drafts of the reports submitted to the U.S.O.E. and their comments have played an important part in the planning of this Report.

The writing of this Report has been a collective effort in the sense that each member of the Abt Associates Follow Through staff contributed in some form to each segment of the Report. Individuals were, of course responsible for writing the final versions of each chapter and it is appropriate to acknowledge these contributions here. Those who were primarily responsible for the writing of each chapter are: Chapter I, Michael Kane; Chapter II, Marvin Cline; Chapter III, Linda Stebbins; Chapter IV, Richard Bale; Chapter V, Madhukar Joshi; Chapter VI, Elizabeth Proper, Nancy Ames and Carolyn Stern; Chapter VII, Richard Anderson, Thomas Ferb, Nancy Ames, Duane Hybertson, Donna Park and Marvin Cline; Chapter VIII, Thomas Ferb, John Larson and Marvin Cline; Chapter IX, Marvin Cline.

Volume IB contains a series of monographs reporting special studies. These were written by Elizabeth Proper (Monograph I), Nancy Ames (Monograph II), Carolyn Stern (Monograph III) and Madhukar Joshi (Monograph IV).

In addition, Micheline Conte served as research assistant on several studies, and Dr. Charles Hall provided technical advice on the statistical interpretation of several of the findings. Linda Stebbins was responsible for managing the editorial processes which resulted in the final version of this Report.

The task of preparing the data for analysis, writing software and running analyses was accomplished by our data processing team including Kenneth Carlson, Paul Freyheit, Sanford Friedman, Michael Hagerty and Nouna Kettanah. This was a particularly massive and complex task that could not have been accomplished without their impressive technical skill and good humor.

Finally, we would like to thank Muriel Kendrix who served as Administrative Assistant to the Project Director and without whom the myriad details of managing the project would have brought the work to a halt at several critical points. Jane Allison, assisted by Ann Czajkowski served as Project secretary with impressive amounts of patience and skill for which we are indeed very grateful and indebted.

Marvin G. Cline
Project Director

CHAPTER I

AN OVERVIEW: THE GROWTH OF THE PLANNED VARIATION MODEL

1.0 INTRODUCTION

The design of the Project Follow Through Planned Variation study is varied, complex, and longitudinal in nature. It is predicated on the assumption that children who attend preschool programs such as Head Start acquire important advantages and that these advantages can be maintained in the public schools with the appropriate enrichment of public education. Although the meaning of appropriate enrichment is not clearly known, it is assumed to include innovations in curriculum, reorganization of school systems, increase in parental involvement in the educational process, and the provision of comprehensive medical, social, psychological, and nutritional services to children.

The emphasis of the Planned Variation experiment is on the "development, refinement, and examination of alternative approaches to the education and development of young disadvantaged children." (Egbert, c. 1971) Twenty-two groups of elementary education specialists (Sponsors) are now working with school districts to test their approaches to the problem of enrichment in the public school setting. A subset of this group of Sponsors was selected to participate in the national evaluation. In this chapter we will describe the origins and nature of the Follow Through (FT) program and the programs of those ten who were included in the analysis summarized in this report.[1] The remainder of this report will examine this program and its patterns of effects on children, teachers, parents, school systems, and communities during the course of the kindergarten year and beyond.

2.0 ORIGINS

An early evaluation of Project Head Start (Wolff and Stein, 1966) indicated that although school readiness was increased by the 1965 Summer Head

---

[1] For a more complete description of Project Follow Through and all 22 Sponsors, see USOE (1973).

Start experience, it was not reflected in achievement test gains at the end of kindergarten in 1966. While some critics saw this study as raising questions about the value of Head Start, the Johnson administration took it as an opportunity to extend a Head Start type program into the public schools by requesting a program to follow through on Head Start gains in the early years of schooling.

On January 10, 1967, Follow Through was formally proposed in President Johnson's State of the Union Message. Under the Economic Opportunity Act he requested 120 million dollars in fiscal 1968 for a Follow Through program for up to 200,000 children.

Before the legislative proposal received congressional approval, officials of the Office of Economic Opportunity (OEO) and the U.S. Office of Education (USOE) began planning a broad-scale program to extend Head Start's comprehensive social and educational services into the primary grades. The method by which Follow Through would eventually be administered emerged from this early planning phase.

Follow Through, authorized under the Economic Opportunity Act, would be administered under a delegation of authority from OEO to the Department of Health, Education, and Welfare (DHEW). Within DHEW, the Division of Compensatory Education, Bureau of Elementary and Secondary Education of USOE would have responsibility for the Follow Through program. The Memorandum of Understanding delegating the program's administration carried two critical points: (1) final authority for the allocation of funds rested with OEO; and (2) projects funded were to include all major components of OEO community action programs. The latter point underscores the fact that Follow Through was intended to extend the Head Start community action model into the public schools. The criteria for funding developed by USOE included the OEO requirements that the projects offer: (1) comprehensive psychological, social, and pupil services completely integrated with classroom activities; (2) maximum use of school and community facilities and resources; and (3) meaningful parent and community participation in the planning, implementation, and operation of the program.

Before Congress passed the legislation authorizing the Follow Through program, OEO advanced 2.8 million dollars to USOE to initiate pilot projects.

These funds were to be returned to OEO out of the first funds Congress appropriated for the program. USOE was enabled to fund planning grants in 40 pilot school districts during the Summer of 1967. Operational grants were made to 30 of them in the Fall of 1967 and ten more school districts were added by the end of the year.

During this time major revisions in the basic nature of the program were underway. It was anticipated that funding for fiscal year 1968 would be at the 120 million dollar level the President had requested. This would permit a greatly expanded program for the 1968-69 school year. This was not to occur. The OEO budget as finally authorized by Congress was one-eighth of that requested. Follow Through, funded in the OEO budget but administered through another agency, became low on the list of OEO priorities. Expecting 120 million dollars, the program received 15 million dollars of which 3.75 million had already been borrowed and spent in the 40 pilot projects. The impact was obvious. Follow Through became a much more limited program than it was originally conceived to be.

Since funding levels made a full-scale service program impossible, it was decided to use the program funds to determine "what works." That is, the new program emphasis was to systematically introduce a variety of well defined programs into the kindergarten through third grade sequence and systematically evaluate the effects of such variation. Although this approach, which came to be known as the Planned Variation model of educational experimentation, was never formalized, it was generally agreed to by officials in the relevant federal agencies (Egbert, 1973). Thus the intent of the program changed from service to experimental, but the authorizing and enabling legislation remained unchanged. This undoubtedly produced a wide variety of problems, the most important of which, from the point of view of the national evaluation of the programs, was to curtail the variables with which the program could experiment.

One outstanding example of a set of variables which was excluded from the national evaluation of the Planned Variation model includes the medical/dental, social service, and community action components of the program. By Congressional authorization, Follow Through is a community action and social service program. The program is mandated to contain the following:

- medical and dental services;
- a nutrition program;
- a social service program;
- guidance and psychological services;
- community and parent involvement including, but not necessarily limited to, a Policy Advisory Committee which must draw over half its members from parents of Follow Through children and play a substantial role in the planning and management of the project; and
- participation of community agencies

The major difficulties in measuring these critical variables could not easily be overcome without some modification of the legislation to reflect the shift in program emphasis. Consequently, these variables were not included in the national evaluation of the Planned Variation model. The experiment was limited instead to the domain of the instructional approaches. Guidelines for participation in the experimental component of the programs included, therefore, the following:

- participate in the Planned Variation experiment including, for most projects, affiliation with a program Sponsor;
- articulate primary programs with preschool programs;
- engage in training and development; and
- provide for the use of paid paraprofessionals and volunteer workers.

3.0 PLANNED VARIATION EXPERIMENTATION THROUGH PROGRAM SPONSORS

In order to operationalize the concept of Planned Variation, USOE developed the notion of educational specialists, each sponsoring a different educational model in a group of school districts. This strategy was novel for two reasons. For the first time research institutions, institutions of higher learning, and others with theoretical experimental notions about the education of children were asked to transfer their ideas from the college classroom, the textbook, or the laboratory school setting into the public school classroom on a large scale. Second, school districts which had previously been totally independent of outside intervention were asked to

enter into a partnership with a change agent. Hence participating school districts were required to select a Sponsor (or an independent educational approach) and work with this Sponsor in the implementation of that approach.

As the concept of Sponsorship was developing, it also became evident that very few well developed approaches to early education of disadvantaged children were ready to implement in the primary grades. In order to get the project underway, however, USOE officials indicated that the Sponsor of each experimental model would be expected to develop and refine the model as experience in the field was acquired. The refinement process included becoming more proficient in implementing the model under a variety of political, social, and educational conditions. In fact, most Sponsors had to develop implementation plans and strategies as well as instructional plans simultaneously. Implementation of Follow Through was a project without precedent.

4.0 PROGRAM IMPLEMENTATION

The selection of school districts to participate in the Follow Through Planned Variation experiment was a complex process. Initially, chief state school officers and state OEO officials were asked to nominate school districts for participation. From administrative necessity, the criteria used for nomination and selection reflected more the difficulties of program administration than the requirements of scientific experimentation and sampling. Of 225 districts nominated, 51 were chosen in mid-January, 1968, as grantees. To these 51 and the 40 original pilot projects, 57 more sites were added in 1969-70 and 12 more in 1970-71. The selection procedures for these additional sites were similar to those used for the original 51.

The selection of Sponsors was straightforward but, once again, not primarily concerned with experimental design. USOE had already identified some potential Sponsors during earlier planning meetings. Further canvassing of the national educational community yielded 18 groups who had developed new approaches to elementary education. Sixteen of these groups responded to a USOE invitation with proposals to serve as Follow Through Sponsors. Fourteen were chosen by the first communities involved. In

1969-70 six more Sponsors entered the program and at a later date still two more were added.

Finally, the process of matching Sponsors with projects was influenced as much by the desire to allow school districts freedom of choice as by experimental considerations. During late February, 1968, the 16 potential Sponsors and "hundreds of representatives of school districts, Head Start programs, Community Action Agencies, parent groups, and state agencies were brought together." (Egbert, c. 1971) Sponsors made presentations on their approaches while community representatives and school officials listened, seeking a Sponsor who seemed compatible with their needs and attitudes. At the conclusion of the conference, Sponsors were selected by one or more of the districts involved. Some districts affiliated with their first choice, others with their second or third. Thirteen of the 40 original pilot districts had exercised their option not to affiliate and become "self-sponsored." The 51 new districts, however, were required to affiliate. Fourteen districts were classified "parent implemented" because their programs were to be developed and run by parents and community organizations.[2]

In sum, it must be realized that, given the practical requirements of administering a large-scale federal program, there was no conscious attempt to randomly select participants from the universe of eligible districts, nor to randomly select educational treatments (Sponsor models) from a universe of possible treatments, nor to assign Sponsors to projects in a random manner. The selection process turned out to be one of relatively free choice of school districts from the options constructed by USOE. This is blending of scientific principles with an open, pluralistic system of education, and one which, if it yields useful experimental results, could be the model for future social experimentation.

---

[2] The "self-sponsored" and "parent implemented" categories refer to the process by which the project is designed and managed rather than to the educational treatment occurring in these projects.

5.0 FOLLOW THROUGH PROGRAM SPONSORS

The essential element of the planned variation aspect of Follow Through is the implementation of a variety of educational approaches by program Sponsors. The Sponsor is the outside change agent responsible for working with individual projects to deliver a new approach to education in the project's classrooms. There are 22 Sponsors working with nearly 170 projects throughout the nation to develop and refine successful approaches to instruction. Although the instructional approaches vary, all Sponsors share common orientations.

- All of them seek to develop children's learning abilities.
- All recognize the importance of individual and small group instruction and frequent exchange between children and concerned adults.
- All are committed to making learning interesting and relevant to the child's cultural background.
- All believe that the child's success in learning is inseparable from his self-esteem, motivation, autonomy, and environmental support. (USOE, 1973)

While all Sponsors are committed to these orientations, the degree of their commitment and their approach to operationalizing it varies widely. So too do the psychological and philosophical bases underlying each Sponsor's approach. Some are more oriented toward academic achievement while others are more concerned with developing a process of instruction which will instill a desire to learn. Still others are oriented to teaching how to learn. Some Sponsors appear to be very similar in approach, others widely diverse. Regardless of appearances, all Sponsors do differ; yet all pursue the educational and social objectives of the Follow Through program.

The concept of planned variation is intended to help determine which of a variety of possible educational approaches works best in which of a variety of settings. The program Sponsor is the basic building block of this effort.

Whereas there are 22 Sponsors in Project Follow Through, only ten of these have been included in the analyses and discussions which follow. While these Sponsors are likely to be representative of the full spectrum of instructional approaches, they were chosen because they have sufficient information in the data base to make

analysis of these data feasible. Each of these ten Sponsors is described below.

5.1 Sponsor 2: Responsive Educational Program Far West Laboratory

Evolving from the belief that a healthy self-concept allows a child to appreciate himself, his culture, and both his abilities and limitations, this model provides the child a learning environment in which he can explore and discover. Within a carefully designed setting the student is free to choose those activities he wishes to engage in. The goal is for this freedom and exploration to result in the child making interrelated discoveries about his physical and social world, all the while developing a healthy self-concept, and knowledge.

This autotelic (self-revealing) approach holds that the best way for a child to learn is for him to be in an environment in which he can try things out, risk, guess, ask questions, and make discoveries without serious psychological consequences. The learning centers, tasks, and games utilized by this model structure such an environment to some extent. The materials and the child's interaction with them are self-rewarding and stimulate the development of self-direction and inner controls. Teachers provide guidance but the child works on his own. There is no set pace. Learning sequences have been developed but each student works at his own rate. The model assumes that no single theory of learning can account for all the modes in which children learn; therefore, it seeks to provide a variety of educational alternatives which build on the background, culture, and life-style the child brings with him to the classroom.

Objectives:

- Make available a variety of education alternatives in the classroom so that the child is free to explore and to set his own learning pace.
- Develop the instructional staff to become more responsive to the individual child's needs.
- Develop the problem solving abilities of the child.
- Help the child develop confidence in his own capacity to succeed.

- Help the child develop the academic skills necessary for effective problem solving.
- Develop a learning environment that helps the child make interrelated discoveries about his physical and social world and develop a healthy self-concept. (USOE, 1973)

5.2 Sponsor 3: Tucson Early Education Model
University of Arizona

This model holds that an educational program should provide a child with a variety of experiences which will develop both his academic and social ability to function effectively and confidently in society. The skills and abilities required for participation in contemporary society are missing in the behavioral repertoires of many individuals because their background does not provide an adequate basis for their development. The educational experiences of the Arizona model seek to overcome this perceived deficit. Skills are always taught in a functional setting, and concepts are illustrated with examples from areas both within and outside the classroom. Teachers individualize and emphasize adult-child interaction on a one-to-one basis. Recognizing the differences in needs and learning rates of children, a great variety of behavioral options of both a self-selected and structured nature are provided students.

The curriculum focuses on four general areas of development: language competence, development of an intellectual base, development of a motivational base, and social arts and skills. The classrooms in which these are elaborated are organized into behavioral settings and interest centers for small groups, to encourage interactions among the child, his environment, and others. In addition, this classroom organization encourages social reinforcement techniques while the curriculum materials used are in themselves arranged for their reinforcing value.

Objectives:

- Develop the child's ability to think as facilitating the learning process.
- Develop the child's social and academic skills toward effective social interaction and communication.
- Develop attitudes and behavioral patterns which will enhance the total learning and socialization process for the child. (USOE, 1973)

5.3 Sponsor 5: Bank Street College of Education Approach
Bank Street College

Bank Street believes that learning of specific skills should not take place independent of healthy emotional development. Learning and development are intertwined. Learning must be pursued by the child on behalf of his own development; if not, it will be superficial. Therefore, the Bank Street classroom is designed to offer a rational and democratic situation in which a child's positive image as a learner and a person can develop.

The classroom is the child's workroom. He participates actively in his own learning as the adults in his room support his autonomy while expanding his world and sensitizing him to the meanings of his experiences in it. Academic skills are acquired within the broad context of planned activities that provide appropriate ways of expressing and organizing children's interests. The classroom is a stable organized environment. The teacher introduces activities and plans events but always in terms of the individual child's response. The teaching is diagnostic with a strong emphasis on individualized follow-up. While the planned activities originate from classroom themes such as organizing chores, or block building, they later extend to community themes (marketing, traffic, and water safety).

In the Bank Street classroom the focus is on tasks that are satisfying in terms of the child's own goals and productive for his cognitive and affective development. Academic skills are learned in a context of a relevant, engaging classroom life.

Objectives:

- Provide an individualized curriculum.
- Enable children not only to acquire basic knowledge and skills but also to master how to learn.
- Encourage communication which is self-initiated, creative, and expressive.
- Develop agreed-upon limits for behavior with full freedom of expression within these limits.
- Create a learning environment to challenge the child and to stimulate and support probing and problem solving.
- Extend the learning experience beyond the walls of the classroom.

- Encourage reinforcing relationships within each teaching team and among all staff members.
- Involve parents in classrooms and in social and community activities related to the school.
- Provide opportunities for parent-staff interaction on behalf of the individual child and the total program. (USOE, 1973)

5.4 Sponsor 7: University of Oregon Engleman/Becker Model for Direct Instruction
University of Oregon

Engleman and Becker believe that children will learn if they are taught well and there is a payoff for learning. They insist that a child who fails is one who has not been taught properly and that the remedy lies in teaching the skills that have not been mastered. The model holds that disadvantaged children can perform at "normal" levels of achievement when the instructional program builds, at an accelerated pace, upon the skills they bring to school. Therefore, the primary concern of this compensatory program is to teach academic skills and teach them rapidly.

In the model's classrooms at least one hour a day is spent on academic skills--reading, arithmetic, and language--in small group situations. The use of reinforcement is a key element in this aspect of the program. Children are smiled at and praised for correct performance. The materials are programmed and sequenced so that the tasks a child encounters are not too difficult. The teacher works with only four to six children in a rapid paced question-answer model; the children respond in unison in a prescribed fashion. In this manner the teacher receives continuous feedback on the performance of children and children are immediately rewarded for good performance.

Objectives:

- Bring the child up to the normal level of achievement by building on the skills which the child brings to school.
- Achieve a faster-than-normal rate of mastery of basic learning skills. (USOE, 1973)

5.5 Sponsor 8: Behavior Analysis Approach
University of Kansas

The behavior analysis model uses systematic reinforcement to encourage desired behavior. A token exchange system is set up to provide precise, positive reinforcement of desired behavior. The tokens provide an immediate reward for successful completion of a task. Later these may be exchanged for an activity that is desired, such as playing with blocks, listening to a story, or recess. The token system prevents the immediate delivery of a reinforcing activity from interfering with the behavior which is being rewarded.

The Kansas program holds that an effective system of reinforcement makes the reward contingent on improved academic or social performance. Yet the token system does not preclude the possibility that learning itself can be rewarding. The tokens are used only to support early efforts in a particular area. As the child achieves a level of mastery where the new skill itself is rewarding, the token reinforcement is decreased or discontinued. The teacher's role is that of a behavior modifier. She can monitor the child's progress by noting the amount of tokens he has available to exchange. Thus, the token system provides feedback to the teacher as well as the child.

The token system is used in both the social and academic areas. Children are reinforced for appropriate student role behavior, as well as for progress in reading, language, writing, and mathematics. Programmed instruction materials are used to allow for individualized instruction and to further facilitate teacher monitoring of rates of progress. The model calls for careful and accurate criteria and instructional objectives and this is made possible in large measure by the programmed instruction approach.

Objectives:

- Facilitate and accelerate the child's mastery of basic skills, particularly in reading and arithmetic, through the establishment of a "token economy" within classrooms.
- Train instructional staff to teach appropriate academic and social skills through the systematic use of positive reinforcement and the elimination of punishment and coercion.

- Train instructional staff in the use of programmed curriculum materials so that each child is enabled to work effectively at his own rate of speed.
- Train parents to work (as paid staff) in the classroom so that they will have the opportunity to influence their children's future education through the use of behavior analysis techniques. (USOE, 1973)

5.6 Sponsor 9: Cognitively Oriented Curriculum Model High/Scope Educational Research Foundation

This model represents a synthesis of research in preschool and early elementary education. It focuses on an "open framework" classroom which combines an emphasis on active experience and involvement of the child; a systematic, consistent, and thoroughly planned approach to child development and instruction by the teacher; and continuous assessment of each child's level of development so that appropriate materials and activities can be provided.

The curriculum is cognitively oriented and takes into account the differences between the way children and adults "think." The model's aim is to develop in children the thinking skills they will need throughout their school years and adult lives. The emphasis is on the process of learning rather than a particular subject matter, although the academic subject competencies traditional to the elementary years are taught. The model is an active one. It holds that learning should be active and that it takes place through the child's action on his environment and his resultant discoveries.

Objectives:

- Nurture in the child the thinking and communication skills he will need throughout his school years and his adult life.
- Develop the child's ability to make decisions about what he is going to do and how he is going to do it.
- Develop the child's ability to express himself--to speak, write, dramatize, and graphically represent his experiences and communicate these experiences to others.
- Develop the child's ability to comprehend others' self-expression by reading their writing and understanding artistic and graphic representation.

- Develop the academic subject competencies through application of developing thinking abilities.
- Develop the child's ability to work with other children and adults, so that work done is a result of group planning and cooperative effort.
- Develop the child's self-discipline, his ability to identify personal goals, and to pursue and complete chosen tasks.
- Help the child develop a spirit of inquiry and openness to knowledge and the points of view of others. (USOE, 1973)

5.7 Sponsor 10: Florida Parent Educational Model University of Florida

This model is based on the premise that it is not enough to change the way the school teaches children; one must also change the way their mothers teach them. Therefore, in this program teaching occurs in both the home and the school and is coordinated by a paid parent educator who comes from the same population as the children's mothers. The parent educator is trained by the project personnel. In the classroom she functions as a teacher's aide, but outside the classroom she instructs mothers in how to teach the child and follow up on his classroom activities. Thus, the mother learns the importance of the home in the child's development and education; she learns what activities to encourage, which to discourage, and perhaps most important, that her actions can have an effect on her child. Mother is encouraged to report to the parent educator which strategies seem to work. She is recruited, therefore, as an active agent in her child's growth and development.

The intrinsic rewards this model yields for the parents are stimulating to the child in that they encourage an environment of pride, achievement, and of high self-esteem. In short, the program seeks to overcome the cycle of despair and low self-concept frequently found in low income populations by encouraging a process of active parent involvement. While the curriculum is not standardized, it does have a Piagetian orientation. The child is encouraged to be experimental rather than repetitious but no particular use of rewards is made. Mastery itself is felt to be its own reward.

Objectives:

- Improve the child's school achievement through work on tasks at home.
- Expand the child's learning environment beyond the school.
- Educate parents to participate directly in the education of their children.
- Motivate parents to develop a home environment that stimulates better performance by the child in school and in life.
- Develop a home-school partnership in all areas of school activities.
- Educate school personnel to support and encourage parental cultural contributions. (USOE, 1973)

5.8 Sponsor 11: EDC Open Education Program
Educational Development Center

This program is derived from the British Infant School approach, which evolved over the past few decades. It also draws heavily on the knowledge gained in child development over the past 50 years. The approach is essentially a program for helping communities generate the resources to implement open education.

EDC believes that learning is facilitated by a child's active participation in the learning process and that a fundamental educational aim is for children to assume responsibility for their own learning. Learning, therefore, takes place best in a setting where there is a range of materials and problems to investigate which complement the range of ways different children learn.

In an "open" classroom there is a rich environment of materials for children to explore. They are encouraged to initiate activities, be self-directing, and become intensely involved in their interests. Typically there is a variety of activities going on, many of them interdisciplinary. Time is flexible and self-management is the norm, yielding an atmosphere of cooperation where children work together and help each other learn. There may be many interest areas in the room, some reflecting traditional subject matter distinctions such as social studies or mathematics. The classroom is characterized by an interaction of subject matter and purposeful mobility and choice on the part of children.

The role of the teacher in the open classroom is an active one. The teacher leads children to extend their own projects through thoughtful responses and suggestions. This responsive, insightful person enters into the child's growth as a guide who is constantly involved, not as a director or spectator. The objective is to get the children involved in things that are relevant to them. To do this all things are potentially legitimate, although reliance on a structured, prepackaged curriculum is discouraged. The content of what is taught is also rather open, being most strongly influenced by local conditions and objectives. In this approach the emphasis is not so much on content but rather on a process. Within the successful open classroom, learning to take responsibility for one's own learning is perhaps the most important goal.

Objectives:

- Create classroom environments which are stimulating and responsive to a child's individual needs and which make full use of the talents and creative styles of the teachers and aides.
- Develop academic skills in flexible, self-directive ways that allow learning to become part of children's life-styles outside as well as in the classroom.
- Provide resources and environment for children's growth in problem solving skills, ability to express themselves creatively in their social and emotional development, and their ability to take responsibility for their own learning. (USOE, 1973)

5.9 Sponsor 12: Individualized Early Learning Program University of Pittsburgh

This program is based on the proposition that if a child is to learn most efficiently he must proceed at his own rate. If a curriculum is to teach most efficiently, components must be carefully, optionally sequenced. The project has, therefore, developed a highly structured and interrelated curriculum. It is based upon a component analysis wherein objectives are stated, requisite skills are specified in behavioral terms, lower level skills are deduced and specified, and higher level skills identified, until a clearly articulated hierarchy has been derived both from logic and a knowledge of psychology. Tests which measure the acquisition of these skills are constructed in a similar manner. These provide a check on the

child's progress, the teacher's success, and the adequacy of the component analysis.

There are three general classes of skills included in the curriculum which are felt to underlie all higher order functioning. These are (1) orienting and attending skills; (2) perceptual motor skills, including gross and fine motor skills, such as visual and auditory perception; and (3) conceptual and linguistic skills which include classification, reasoning, memory, language, and early mathematical concepts.

Children learn, the model assumes, by interacting with materials and other children. The teacher serves as a facilitator, monitor, and reinforcer. The teacher, using the diagnostic tests, helps the child move along through the component curriculum using the least powerful reinforcers needed until the child is able to work independently of reinforcement.

Objectives:

- Identify each child's strengths and weaknesses and provide the child with a personal program of instruction based on his individual needs. (USOE, 1973)

5.10 Sponsor 14: Language Development (Bilingual) Education Approach Southwest Educational Development Laboratory

This approach is a design for classrooms where a majority of the pupils are Spanish-speaking. The model holds that language is the child's main tool for dealing with his environment, expressing feelings, and acquiring skills, including nonlinguistic ones. An underlying premise is that learning in a second language is easier and more effective if the child first learns concepts and content in his native language and if intensive oral language development in both languages precedes the learning of literary skills. In addition, a positive emphasis on the child's native language and culture is essential to the development of a positive self-concept and pride of heritage.

Step-by-step sequential procedures are followed in teaching language patterns. Both teaching procedures and materials are designed to develop a hierarchy of thinking processes. The focus in teaching language is on content such that all classroom activities reinforce language development.

The model stresses a high level of adult-child contact. Teachers and aides are constant language models giving the child frequent assurance and reinforcement. The kindergarten class, which stresses visual, auditory, and motor skills, as well as thinking, discovery, and English language structures, is divided into small groups which work both independently or with a teacher. In the first and second grades, where oral communication as well as reading and writing skills are stressed, the teacher presents a lesson to the whole group and then the children work independently in small groups.

Objectives:

- Train instructional staff to appreciate the child's culture, to act as good language models, and to become proficient in language development activities.
- Utilize the child's existing concepts as a basis for sequential development of more advanced concepts.
- Teach children to understand, listen, speak, read, and write with equal competence and facility in both the native language and English. (USOE, 1973)

CHAPTER II
GOALS OF THE INTERIM ANALYSES

1.0 INTRODUCTION

The major goal of these interim analyses is to assess the overall effects of Follow Through (FT) upon the outcomes measured by the battery of instruments administered to the kindergarten children of Cohort III (1971-1972). Given the diversity of the Sponsors' objectives, approaches, and site-specific circumstances, we expect to find a highly diverse set of patterns of Sponsor outcomes. At the end of kindergarten, each Sponsor should begin to show a pattern of outcomes which reflects the impact of the program on the particular kinds of children with whom the Sponsor is involved, under the unique conditions of program administration. In these analyses, we begin to see the first signs of the different effects produced by each Sponsor on the kinds of children in the kinds of localities, involving the kinds of school systems which constitute the real world for that Sponsor.

We approach this major goal in a number of ways. In each approach we contrast the FT children associated with each Sponsor with the corresponding non-Follow Through (NFT) comparison children who were selected by the National Follow Through Office. Since the effects of Follow Through are inevitably confounded with a variety of extraneous factors, we introduce appropriate adjustments into the analyses where possible. Where we cannot adjust, we provide the necessary warnings so that the reader will understand what competing hypotheses we have not been able to eliminate.

Since no single analysis completely settles the substantive question that motivated it, we often approach a given question in several complementary ways and add descriptive summaries of its context. This strategy has yielded analyses which address the following related questions:

- What are the FT/NFT contrasts in posttest scores on achievement, motivation, and absence measures for each Sponsor's Cohort III kindergarten groups, and for Follow Through overall?
- What are the FT/NFT contrasts in posttest scores for each Sponsor implementing a program in New York, Philadelphia, and Chicago?

- How do preschool experience, initial achievement level, sex of the children, ethnic background of children, and classroom integration influence the various FT/NFT contrasts?
- To what extent can we attribute observed FT/NFT contrasts to the unique curriculum inputs prescribed by each Sponsor's model?
- What kinds of teachers are delivering the various FT models?
- From what kinds of home environments do FT and NFT children come? What are their parents like?
- What notable problems have the Sponsors encountered in implementing their models?

To make clear our reasons for selecting these questions, and to set forth the context in which our results should be understood, we now present some details of the substantive, operational, and analytic limitations which characterize these interim analyses. With the limitations in mind, we then summarize the analytical procedures we have followed.

2.0 LIMITATIONS AND QUALIFICATIONS

Some of the practical limitations of these interim results are attributable to the nature of the Follow Through quasi-experiment. Other limitations follow from the design of the experiment, others from the interactions between the design and the real world, and still others from the constraints imposed by the analytic procedures. In order to clarify our reasons for undertaking the specific analyses that make up this report and not others that our substantive concerns would seem to suggest, we now discuss the principal categories of constraints that circumstances impose on these analyses.

2.1 Prematurity

Although our findings to date generate a number of significant evaluative statements about Follow Through and its Sponsors, this is only an interim report. It is not intended to present definitive answers to any of the questions that motivate it. For some categories of questions, indeed, we have as yet very little to say.

Follow Through to date has provided Sponsors with a four-year opportunity to develop approaches to implementing their models. The diverse curriculum models participating in the program can be expected to have different kinds of impacts on children at various times during

the period from kindergarten to third grade. Some Sponsors expect immediate effects, because they focus on the acquisition of traditional skills from the first day of contact with kindergarteners. Other Sponsors are oriented toward problem-solving behavior which might not yield an immediately obvious impact on the acquisition of traditional skills. Still other Sponsors emphasize the stimulation of particular developmental processes in the cognitive domain; others are concerned with affective processes. Both the time at which effects are to be expected, and the kind of effects which each of the models might expect, vary from Sponsor to Sponsor. These programs are designed to be significant alternatives to traditional programs of primary education involving changes in school and classroom organization, teacher training programs, teacher attitudes, and parental and community involvement in the educational process, as well as curriculum changes. The developers of Follow Through have generally considered that the smallest time span in which Sponsor impact on children, schools, teachers, and communities, can be expected to become observable is three or four years.

The data available for these analyses do not include any four-year data. Next year's data will give us the first opportunity to look at four-year results. Even then, we shall have the four-year longitudinal data only for Cohort I, where it will be partially obscured by the confounding influences that have beset the implementation of Follow Through. The Sponsors' efforts with the first two cohorts involve trial, planning, and unanticipated problems. Cohort I was the first group of children with whom most of the Sponsors applied their models in anything more than an experimental exploratory form. Each successive grade that these children entered represented an entirely new experience for the children, their teachers (in most instances), and the Sponsors. Cohort II entered at a time of expansion to new schools, new teachers, and new communities. Cohort III children, on the other hand, entered the programs when the Sponsors were relatively experienced as innovators, change agents, or teacher trainers (or any combination of these depending upon the Sponsor). Cohort III, furthermore, is by far the most heavily sampled in kindergarten (N_k = 19,841; N_{tot} = 26,567) of the three Cohorts. Combining

kindergarten and entering first, FT and NFT, 2,530 children were tested in Cohort I and 22,576 children were tested in Cohort II. A smaller group of Sponsors ($N_1 = 14$, $N_2 = 20$) of alternative models of primary education were included. Cohort III thus represents the first opportunity to follow a sufficiently large sample of schools and children over an extended time period to test adequately the notion of Planned Variation.

In this report, therefore, we focus primarily on data from the tests administered during the kindergarten year to approximately 10,000 Cohort III children associated with ten Sponsors either as Follow Through (FT) program participants or as non-Follow Through (NFT) comparison subjects.

These children entered kindergarten in the Fall of 1971. At this writing, they have completed their first grade year; new test scores are now being prepared for next year's analyses. These children will continue to participate in the Follow Through study until the Spring of 1975, when most will have completed the third grade. At that time, they will be given an end-of-program battery of tests so that we may assess the full four-year impact of the Follow Through programs. At present, this Cohort III population affords our first substantial opportunity to look at one-year effects of Follow Through. We anticipate, furthermore, that the relatively heavy initial sampling of Cohort III will allow a better sample of third grade "survivors" for the final assessments than will be available for the earlier Cohorts.

Despite their drawbacks, we shall use data from the first two Cohorts in several future analyses in which we shall examine the impact of selected Sponsors over successive years in the same grade or in the same schools. The three-year longitudinal study and the multiple-cohort study of Chapter VII - 5.0 illustrate the forms that these analyses will take when the necessary data become available.

The diversity of Sponsor objectives suggests another analytic dimension that we shall investigate when the data base has developed somewhat further. The concept of Planned Variation suggests that each

Sponsor should produce a unique pattern of effects upon the affective and cognitive measures over a four-year time span. Some Sponsors, for example, expect that cognitive development will be enhanced in those children in whom they have stimulated significant increases in the sense of self-competence, the sense of self-control over the environment, and the motivation to persist in academic behaviors. Other Sponsors, on the other hand, expect that those children who experience carefully nurtured successes in specific academic activities will thereby acquire a sense of competence and heightened motivation. To discern clearly the effects of particular Sponsors with particular children, therefore, we must examine patterns of multiple outcomes. To date, our analyses have been univariate; they therefore do not yet address the critical issue of multiple outcomes. When the second set of data on Cohort III becomes available, we shall undertake multivariate analyses and report on them next year.

Finally, we cannot as yet take account analytically of the many ways in which Sponsors' actual interventions, including their strategies for institutional change in the educational process, deviate from the operational versions of their developmental, learning, and instructional theories and intentions. The original design of the Follow Through evaluation did not incorporate measurements that would permit the unconfounding of dimensions of the Sponsor's "model" and "program."

2.2 Limitations of the Follow Through Design

Experimental designs in general limit the range of inferences and conclusions that their results can justify. For this reason we present both the data summaries of the results and limitations and qualifications which help the reader to attach to each conclusion the appropriate degree of credence. Three major aspects of the Follow Through design which limit the generality and certainty of any inferences from the Follow Through evaluative data are imbalance, purposive selection of subjects, and the qualitative diversity of the Sponsors.

2.2.1 Imbalance

To answer the policy and research questions that motivated Follow Through, the evaluative design should have approximately equal numbers of probabilistically-selected subjects allocated to the FT/NFT groups to be

compared. To the extent that this was not done it is difficult to identify the unique effects of the various factors that define them. In extreme cases, for example, where a Sponsor has no West Coast subjects, or where a site's NFT group includes no urban children or ethnic minorities, the effects become confounded with regional or other extraneous factors.

The Follow Through design achieves balance far better in some respects than in others. FT and NFT groups, for example, are generally of comparable size, making contrasts possible on that basis. The regional distribution of Follow Through sites is far less satisfactory, however, especially within some Sponsors. Chapter III-2.1 describes the extent of the imbalance in regions and also in city size. Some Sponsors are not represented in some sections of the country, and so we can neither examine fully the effects of Sponsors by region nor separate Sponsor effects from regional variations. This imbalance is more than just unfortunate; in some cases, it has kept us from examining some very significant questions: (1) the impact of FT on the full range of ethnic groups in the population, (2) the role of integrated classes for some Sponsors, (3) the contrast of metropolitan sites with smaller sites, and (4) a comparison of programs across the several major regions of the country. Ethnicity, integration, city size, and region are all associated with pupil effects in some way or another, but none is uniformly distributed among Sponsors. One of the most dramatic of these variables is city size. We have found, for example, that Sponsors function differentially within and outside the Big Cities. Different Sponsors, furthermore, have responded differently to this urban challenge.

In presenting our results, in later chapters, we point out possible imbalance effects as they arise. Most of them, unfortunately, cannot be isolated.

2.2.2 Purposive Selection

In a true experiment, subjects are selected from a larger population and assigned to treatment groups probabilistically, so that each member of the population to which one wishes to generalize has a known probability of becoming part of the experimental sample. Follow Through deviates from this ideal in many respects, and so we must make adjustments.

In particular, our analytical subsets (we hesitate to call them samples) are not representative of any definable larger populations. Even though we use probabilistic statistics, we cannot generalize beyond the properties of the groups of children, parents, teachers, or institutions included in our analytic subsets.

This lack of a probabilistic sample has led to numerous unanswered questions. For example, because sites were allowed to select Sponsors, we cannot estimate whether the outcome of a Sponsor's program would be similar to another Sponsor's nor whether other sites would respond similarly to a particular Sponsor. Sponsor-site interactions are confounded with Sponsor effects, and the data do not contain the information we would need to separate the two.

Not only do Sponsors face different problems, but even within Sponsors the FT and NFT groups lack equivalence. In most every instance, schools in which the Sponsors were carrying out their programs (FT schools) were "matched" judgmentally with NFT schools in the same district servicing children from the same kinds of families. Within these comparison schools, kindergarten classes were selected as comparison classes for the local FT kindergarten classes. The match between FT and NFT classes on several relevant domains varies tremendously across schools and Sponsors. This topic will be explored in greater depth elsewhere in this report but it should be known, at this point, that very severe mismatches exist throughout the sample. The major consequence of this fact is that the contrasts between each Sponsor's FT group and the associated NFT group, which are the basis of conclusions drawn about the pattern of outcomes, do not necessarily carry the same meaning for all schools of a given Sponsor. They certainly do not carry the same meaning across all Sponsors. A contrast between, for example, the FT/NFT

kindergarten classes located in rural southern communities and associated with one Sponsor, cannot directly be compared to the contrasts between another Sponsor's FT and NFT groups located in the metropolitan North.

Because NFT comparison groups were defined judgmentally, statistics cannot tell us how large an FT/NFT contrast must be before we can attribute it to the effect of a Sponsor's program. FT and NFT groups start out unequal before treatment, and we must do what we can to take account of their initial inequalities before we can know what to make of their final differences. Clearly, we must make extensive adjustments to the data in order to generate Sponsor patterns that make sense both within and among the several Sponsors.

FT/NFT mismatch arises from a number of sources. Program guidelines specify, for example, what children are eligible for FT states: children of the poor, graduates of Head Start, present in definable concentrations within schools, are the mandated recipients of the program. Those who do not get the program are by definition different from those who are in the treatment groups along dimensions closely associated with both treatment and outcome measures. The guidelines make it likely that FT children will come from lower income groups and will achieve initially at a lower level than NFT children. The early designers of the planned variation study worked hard to build into the guidelines an opportunity to involve a comparison group at each site which closely resembled the FT group on as many dimensions as possible. Monograph III describes the events which surrounded the matching of schools to Sponsors in several sites, and it is quite clear that a myriad of social and political forces external to the FT program dominated the assignment process. As a result of these forces, many schools were assigned to FT status for locally important reasons, and those schools which were available for assignment to comparison status were available for locally important reasons. At some sites, the schools left over for comparison were much higher in socioeconomic status than the FT schools simply because all the low income schools in the area were incorporated into the FT group. At some sites the NFT schools had to be found in adjacent communities which were not eligible for FT participation, thereby separating the treatment group from the comparison group by both geographical and income differences. At some

sites, moreover, local administrators assigned low income children to a single school in order to satisfy requirements for participation. The consequence of this confused assignment process was that in very few cases did the children of the treatment group match the children of the comparison group on all important dimensions. The most important of these is revealed when the tables describing the pretest scores of the treatment group and comparison groups are examined. For almost every Sponsor, the FT/NFT group differences are very clear. Furthermore, they are not always in the same direction. At least two Sponsors show extreme pretest differences between their FT/NFT groups: in one case the FT is superior, and in the other case the NFT is superior.[1]

2.2.3 Sponsor Diversity

Follow Through Sponsors generally focus their models toward major redirections of elementary education, not simply at the introduction of new processes to the old goals. In some cases, they are introducing wholly different sequences of materials from those used in traditional classes. In other cases, they are stimulating wholly new (to the world of elementary education) functions and skills involving exploration, inquiry, self-directedness, and problem solving. The materials and procedures in which teachers and in some cases, parents, are being trained vary greatly in their resemblance to traditional materials, procedures, and sequences. We shall discuss later the consequences of Sponsor diversity and innovativeness for the measurement process. Let us point out here that this feature of the Follow Through design makes

---

[1] In the case of the Sponsor whose FT group is superior on the pretest, this difference may reflect in part the treatment delivered by that Sponsor during the four to six weeks between the beginning of the school year and the time of pretesting. If the treatment effects were entirely responsible for the pretest differences, on the other hand, one would expect much larger posttest differences than actually occur. Consequently, we suspect that the FT/NFT differences at pretest result from both selection procedures and also treatment effects. On the other hand, it is implausible that the second Sponsor's treatment produced the dramatic FT pretest disadvantage that appears. It is clear from Chapter VII that sufficient causes arise at the time of site assignment to Sponsor to account for these differences. Once again, the most compelling hypothesis is that pretest differences result primarily from the site selection procedure.

impractical the identification of the "best" Sponsor. There are simply too many ways in which the Sponsors differ qualitatively to permit any simplistic rank-ordering of Sponsors along a scale of "goodness."

Instead of looking for the best, we seek to investigate what kinds of effects are found with what kinds of children, at what points in time, under what particular conditions of program adminsitration associated with each Sponsor.

2.3 Practical Limitations

Not only does the design of the Follow Through project limit the range of appropriate interpretations of our results, but the Follow Through Sponsor's theories must be implemented and evaluated in a real world, one that does not always correspond to and enhance those ideas. Classroom realities reflect with varying fidelity the Sponsors' intentions, and evaluation methodology can not always measure and lead to an interpretation of all the relevant nuances.

2.3.1 Implementation

How well do the real results of Sponsor intervention reflect Sponsor intentions? We have already mentioned the distinction between Sponsor model and program, but some additional detail will facilitate further discussion.

The model, as we use the term, is the operational version of the Sponsor's developmental, learning, and instructional theories. In most cases the model can be exhaustively described by reference to desired events in the classroom. In several cases, however, the Sponsor designs these desired events to take place in the home of the child. In other cases, these events take place in the policy-making councils of the educational establishment where parents, encouraged and facilitated by the Sponsor, take an active and effective role in the planning of their children's education. In still other cases, some of the critical variables defining a model are those social systems within schools which support the independent and creative behavior of both teachers and children. In all cases, however, we mean by the model those events which the Sponsor assumes lead directly to experiences which facilitate the child's growth.

By program, on the other hand, we mean the strategies that the Sponsor develops to accomplish the institutional changes required for the full accomplishment of the model. These include changes in the relationship of the teachers to the decision making process in the school, changes in the in-service training programs, changes in the teacher selection program, changes in the role of parents and community in the decision making process, and changes in the attitudes

of school personnel toward the value system of the model. When an external change agent such as a Sponsor enters a school district, there is bound to be a variety of responses from the several concerned actors in the district. In some cases, principals have seen Sponsors as a threat to their control of the school. Elsewhere, some principals have seen FT as a means of achieving high status in the system. Teachers have resisted the models in some cases because they disagree with the educational values involved or because they resent the implication that they are in need of professional training. In a number of instances teachers have been at odds with the principal as to the value of the model, and the teacher has most often been the one to back down. Some teachers have been resented by their colleagues because of their selection for participation in the "special program," and the FT teacher's status in the school has been seriously compromised. In some cases, the community has felt a strong affinity to the model and the Sponsor's representatives, thereby supplying a supportive environment to both the trainers and the teachers. In other cases, the community felt that the model was forced upon it and has resented Follow Through from the beginning. There have been a number of instances in which the school administration saw FT as a dumping ground for the most difficult problem children (including the physically and emotionally handicapped). In other instances, FT classes were seen as so enriched that only the highest achieving children could benefit from enrollment. Clearly, a multitude of agenda have operated for the many actors in a school district who must interact with each Sponsor.

It is also true that Sponsors vary considerably in their strategies and skill in dealing with these problems. It is likely to be true that those Sponsors who have been more successful in negotiating change (and this is intrinsically easier among those Sponsors who do not require systemic change to institute their model than for those who require major changes) are also more successful in facilitating the growth of the children with whom they work. That is, the <u>program</u> of each Sponsor can be a source of pupil performance variance. From an analytic point of view this possibility represents a factor which might be confounded with the <u>model</u> as a source of performance.

In order to deal with this issue one must measure both program and model in each Sponsor and study the two factors separately and as they interact with each other. At this point in the evaluative design we cannot unconfound model and program: the original evaluation design did not incorporate the necessary measures. We are beginning an effort in this direction, however, and we expect to be able to separate these factors to some degree in the future. In the meantime, it would be inappropriate to attribute Sponsor effects exclusively to the educational content of the model. Rather, it is wiser to assume that the effects noted for any given Sponsor represent the consequences of a team of specialists trying to deal with a variety of communities, located in various geographical regions, each of whom has a unique attitude toward the Sponsor. The goal of the Sponsor is to establish the most supportive environment possible for the application of the model; the Sponsor's input includes both the various strategies adopted at each site and also the locally influenced character of the model.

2.3.2 Methodological Limitations

Even if the Follow Through evaluation data contained all the information necessary to answer all the questions that motivated the experiment, the current state of the analytic art would still doubtless introduce distortions and uncertainties of its own. Modern educational research simply does not yet know how to measure all the variables that an evaluation of Follow Through should take into account, and despite recent advances, available analysis methods still assume much more orderly data sets than the real world of education produces.

Since the academic tests used in the Follow Through battery are designed to measure the outcomes of traditional curricula (these tests are, in fact, frequently the source of curricula as well as the measure of outcomes), they are hardly the ideal means to assess outcomes of non-traditional programs. Some Sponsors have gone so far as to assert that if their children are doing well in traditional measures of traditional curricula, their teachers might not have been applying the innovative programs generated by the model. The battery of measures may not detect many other important outcomes of the programs.

Our attempts to assess the interactions between motivational and academic scores cannot reveal the true relations between changing senses of self and the growth of cognitive skills if measures of such skills are not in the battery. Much the same can be said for the attempt to measure the relations between the growth of cognitive skills and academic achievement. This is a critical issue in part because it is unreasonable to assess the outcomes of a program with a measure unrelated to the goals and procedures of the program.

The incompleteness of the battery is critical in another sense, however. Until we have tested the causal logic of the educational model we cannot fully understand the relations between the inputs and the outcome for that model. A model may assert, for example, that if children are taught by a teacher who appreciates and is skilled in certain procedures, then the children will acquire certain skills and understandings. If those skills are acquired, then some children, under certain circumstances, will be able to apply them to the academic materials of the classroom, and the knowledge thereby acquired will generalize to the testing situation. At each step, the analysis must follow the logic of the model in order to test its efficacy in producing academic outcomes. If the analysis must skip any of the steps on the way, then the model is not fully tested. This may occur because of factors entirely external to the model itself (a large number of political and non-educational factors may contribute to the failure of any step to materialize), and therefore preclude reasonable tests of the model. In addition, some aspects of the model are difficult to accomplish: this needs to be known in order to improve application as well as to increase the meaningfulness of the test of its efficacy.

The failure to include appropriate measures of each skill which each Sponsor attempts to stimulate[2] clearly precludes testing completely the logic of each model. The Sponsors have made this point many times, and it must be acknowledged at the outset of this report that this problem imposes a severe restriction on the understanding of many Sponsors' impacts. We shall not be able to identify the "best" Sponsor: the criteria

[2] Attempts, early in the history of FT, were apparently made to generate a Sponsor-specific set of measuring instruments. Test development activities were too expensive and time consuming to allow this effort to come to acceptable levels of fruition.

for a "best" model of elementary education can hardly be limited to two standardized achievement measures, three motivational measures and a count of the number of days absent for each child. On the other hand, this battery does provide enough variation in the range of outcomes to allow for a reasonably close examination of the pattern of effects for each Sponsor, and ultimately a contrast of these patterns across all Sponsors; our analysis proceeds within this framework.

The analysis of covariance (ANCOVA) which analyzes the contrasts we report, permits adjustment of observed contrasts taking into account the confounding effects of initial mismatch between contrasted groups and of other variables (covariables) which correlate with mismatch. ANCOVA does help to reduce known and measured spurious influences, but it does not eliminate unmeasured confoundings. Even among the measured covariates, moreover, biases exist to the extent that the covariates are measured imperfectly. We have made use of adjustment techniques that take account of the fallibility of one covariable at a time, but those techniques do not permit us to adjust simultaneously a number of fallible covariables.

Another problem emerges when the logic of this adjustment is considered. Partialling pretest out of posttest scores for an estimate of true post scores yields interpretable results if we assumed equal comparison groups on pretest and on posttest. This is not an acceptable assumption, however, since the groups which are higher initially have their advantage for plausible reasons. The higher group can be assumed to be acquiring score points at a greater rate than the lower group; if this differential rate persists, it will lead to a magnified difference at posttest. This phenomenon, which Campbell (1971) has called "fan spread," suggests that the appropriate baseline for the comparison of the two groups is not the initial differences but the relationship between the differences at pretests and posttests. At present there are no fully accepted solutions to this problem and none were attempted in this study. Our adjustments are certainly accomplishing less than they should. Groups which are moderately apart initially can be expected to be even further apart at a later date. Treatment effects may therefore be present even when no differences are observed in the

comparison of true scores. This makes interpretation of treatment effects rather difficult, and requires that the reader keep in mind that the effects we report here are likely to be conservative if the FT group starts out lower than the NFT group, or exaggerated if the FT group starts out higher than the NFT group.

A major issue remains with respect to fan spread when Sponsor-to-Sponsor comparisons are made. The larger the initial differences between treatment and comparison groups, the less efficient is the covariable in adjusting for these effects. A Sponsor who shows very large initial differences will show larger adjusted posttest differences than a Sponsor with smaller initial differences, even when the treatment effects are essentially the same in both Sponsors. In the most extreme case, it will be very difficult to interpret a comparison of the effects of one Sponsor whose FT group is initially much higher than the NFT group with a Sponsor whose FT group is initially much lower. In this case the Sponsor with the higher FT initial scores will show inflated posttest differences in favor of FT (fan spread will spuriously augment treatment effects). The Sponsor with lower FT initial scores, on the other hand, will show inflated posttest differences in favor of the NFT group (fan spread will spuriously diminish the apparent effects). In the latter case the FT group may in fact be responding quite favorably to the treatment but the apparent differences between FT and NFT may be larger: FT groups may look as if they are falling further behind. Of course, the smaller the FT/NFT initial differences, the more directly we can interpret the adjusted effects. Given the lack of random assignment of subjects to treatments, and the corruption of the assignment process by local political processes, it is important to examine this issue even over a one year kindergarten period.

The details of these methodological issues will be discussed fully in the rest of the report; we mention them here so that the reader will be aware of the full range of the limitations of our findings.

3.0 THE ANALYSIS STRATEGY

We have designed a sequence of analyses which gives the best answer we can now provide to each of our principal evaluative questions, given the constraints we have outlined. We have aimed to identify a number of specific contrasts and to describe several significant aspects of the context of Follow Through. The following sections describe our approach to each of seven goals.

3.1 GOAL 1: To Identify the FT/NFT Contrasts in Posttest Scores on Achievement, Motivation, and Absence Measures for Each Sponsor and Overall.

We can investigate these contrasts at three levels of analysis: (1) child level, contrasting all FT children to all NFT children for a given Sponsor or for Follow Through as a whole; (2) class level, using class means of child characteristics as class characteristics, and (3) school level, using the means of all children tested in the various schools. Each of these approaches represents a rather different kind of contrast, and each addresses a different set of questions. We present in this report results of analyses at all three levels of analysis. We also present data which imply an overall FT/NFT contrast, ignoring Sponsor distinctions. It is very hard to interpret this overall contrast precisely, since Follow Through combines a number of very diverse phenomena. It does provide an overall picture of the effect of Follow Through across the nation, nevertheless, and so we report it.

For the purposes of the first FT/NFT contrasts for each Sponsor, we have chosen to focus on the school level of analysis. The FT treatment is administered at the school level in the sense that a school is first selected for participation in the program, and then all of the eligible or desired classes are selected from within that school. Classes and children are not selected independently of schools; since the treatment is applied at school level, it is therefore appropriate to select the school as the unit of analysis. The scores of all kindergarten children tested in a given school have been summed and divided by the number of children to arrive at a school score. Although the number of children varies from one school to the next, we have reason to believe that each school score is acceptably stable. The analyses executed at the school level are those upon which we will base our interpretations

about the effects of FT within and across Sponsors. These FT/NFT contrasts are of course school contrasts, not contrasts among children.

Although children and classes were not selected independently of schools, it can also be argued that treatments are in large part defined by the behaviors of teachers, only a part of which is a function of the schools in which they teach. It is reasonable to consider the class as the unit of analysis if we remember that class effects are partially confounded with those of the school in which the class is located. For a variety of reasons, class level analyses were based upon a subset of the pupils included in the school level analyses: it will therefore be necessary to consider the sampling biases generated at class level when comparing school and class level main effects. Several variables change their meanings, moreover, when they are aggregated to the level of the class instead of the level of the school. These changes are not just statistical in character. There are some changes in the conceptual meanings of the variables as well: we describe these to help the reader interpret these analyses appropriately.

The samples for the child and class level analyses are essentially the same. At child level again, however, some variables change their conceptual meanings from their meanings at the school level and these changes must be kept in mind when the child level analyses are interpreted.

It is important to keep in mind that the primary analyses of FT/NFT contrasts are those carried out at school level. Class and child level studies are not true replications, since much smaller samples are involved, and there are some important statistical and conceptual differences in some of the variables. Nevertheless, these secondary analyses are reported to add depth to our study of the FT/NFT contrasts. The results of the analyses of both Goals 1 and 2 are reported in Chapter VII.

3.2 GOAL 2: To Identify the FT/NFT Contrasts in Posttest Scores for Each Sponsor Associated With a Sample of Children in Three Large Cities (New York, Philadelphia, and Chicago).

In an effort to reduce the confounding effects of regional variations, a population density and uneven distribution of the analytic subset across Sponsors, the National Follow Through Office decided to concentrate several Sponsors in the same geographical region. Seven Sponsors were then located in these three cities such that their samples and the NFT groups were to be highly matched. To date, the Big City analyses are primarily school level analyses, although there are not enough schools within

this special sample to apply statistical tests with a degree of power necessary to note the signals embedded in the amount of noise present in a study of this kind. Consequently, we have simply examined the total school sample with the Big City sample included and excluded, and we have noted the differences in effects under those conditions.

3.3 GOAL 3: To Identify the Influence of Several Relevant Variables on FT/NFT Contrasts in Posttest Scores.

The variables selected for this report are the preschool experience of each child entering kindergarten, the level of academic achievement of the class at the beginning of the kindergarten year, the ethnic background of each child, the sex of each child, and the ethnic mix of the classroom.

Each of these variables has been selected because it has theoretical interest vis-à-vis developmental issues: each addresses the specific details of the impacts of individual Sponsors, and each has a high degree of policy relevance.

Preschool experience of children is a variable of very great interest to both policymakers and developmental theorists since the rationale for the FT programs is that they will both maintain and build upon the advantages generated in preschool. Our major concern is not in estimating the advantages generated by preschool since we have little data on the kinds of preschools which these data represent (future work will examine the more precisely defined Head Start Planned Variation programs as contributors to the FT effects). We are only interested in knowing how variation in preschool experience is related to the nature of impacts on achievement and motivational measures for each Sponsor. Some Sponsors may be able to build immediately on the preschool experiences but not to move children who have had little or no such experience. Other Sponsors may have positive effects on children with preschool experiences only in the latter grades of the program and not during the earlier grades. If either the immediately obvious, or the subtle "sleeper effects" of preschool experiences are not taken into account in assessing Sponsor effects, the true strengths of the programs may not become apparent. We investigate the effects at child level.

The sex of the child has been shown in the research literature to be a critical factor contributing to the developmental processes of the child. For our purposes, it is clear that girls tend to develop faster than boys in some domains in the primary grades. It is not clear, however, whether these differences are inherent or whether they reflect the differential role expectations that are placed upon boys and girls. We include this variable in order to make sure that we have included in our analytic groups children who are generally similar to those described in the research literature. At the same time, differences in developmental rates between girls and boys suggests that the various Sponsor inputs may have differential effects depending upon the sex of the child. If this is the case, it is desirable both to study these effects, and to utilize sex as an adjusting variable in comparing the effects among Sponsors.

Much the same can be said for the inclusion of the entry level of achievement as a variable which might influence Sponsor effects. High achieving classes present very different conditions to teachers and Sponsors than low achieving classes, and can very easily define the conditions under which Sponsors produce their effects. At class level, we must study this variable cross-sectionally, since classes do not remain intact from year to year and do not retain their entry level status. It is particularly important to examine this variable during the kindergarten year before it is contaminated with treatment effects. It is also particularly important to look at the relative effects in achievement and motivational domains of Sponsors working with classes at different entry levels. One might expect to find relatively greater improvement in motivational domains in the lower achieving classes than in the higher achieving classes, and as the data accumulate over the next years, we shall be able to examine this kind of issue. For now, it is important to note how the Sponsors' programs interact with this variable. Ultimately, we may be able to make precise comparisons among models only to the extent that we understand the unique strengths of each Sponsor with respect to this variable.

The last two variables to be examined in their interactions with Sponsors, are closely linked. They are the ethnic background of the child, and the ethnic mix of the classroom. Because of the very small sample of Chicano, Indian, and other ethnic groups in the national sample, all ethnic groups except Black and White children have been excluded from these analyses. No useful purpose can be served in making any comparisons between

the Black and White children since there is no way of identifying the sources of any differences in motivational and achievement measures which might emerge. The ways in which these children differ in their motivational approaches to school, and their responses to the school's [illegible] to them [illegible] too intimately tied to diffe[illegible] of their [illegible] experiences to account easily for any differences in their scores on these measures. The fact that Black and White children may receive different social and educational experiences because of their ethnic membership (over and above their social class membership), however, may be of paramount importance to the educational community. The current controversy over bussing children to achieve integrated classrooms attests to the possibility that a White antipathy to contact with Blacks remains unchanged in some sections of the nation. The educational community has a clear responsibility to deal directly with both those who demean and those who are demeaned by this kind of social injustice. Educational programs must be developed which are effective not only for lower income children generally, but also for those Black children who have experienced the unique insult of ethnic injustice. Consequently, we must examine the full range of FT programs for those conditions which maximize benefits for each of the ethnic groups. It would be ironic indeed if one of the programs showed a strong effect on the motivational or achievement scores of Black children, but not for the White children, and was judged an inadequate model because in the aggregate no meaningful effects could be observed. And, in fact, such a situation does prevail in the case of one Sponsor.

One other reason for studying the interaction between ethnicity and Sponsor effects (i.e., the conditions under which effects are maximized for each group) is that if the programs are effective, it would be expected that these interactions would be significant in kindergarten and first grade but non-significant in the latter grades. If the FT programs are achieving the equality of opportunity which is their mandate, then the patterns of effects which are uniquely associated with the Ethnicity X Sponsor interactions would ultimately disappear. If the degree of equality of educational opportunity can be assessed by the decreasing size of the correlation between ethnicity/social class, and educational outcomes, then it is necessary

to examine these correlations over time to estimate the increase in equality of opportunity. Clearly it is still psychologically significant to be either a Black or White child in the public schools of today, and it is necessary both for the assessment of individual programs and for charting the reduction in that significance, to include ethnicity in the analyses of Sponsor effects.

Ethnic mix of classroom is also a critical variable on which Sponsor effects must be examined. Black and White children are in many cases still able to be enrolled in the same classes and this is a fact of very great educational significance. The atmosphere of integrated classes, the responses of teachers, and the interactions of Black and White parents living in the same communities all may have a meaningful impact on the motivational and achievement scores of children. Some Sponsors may be uniquely capable of capitalizing on these factors, and this would be a finding of major note. Other Sponsors may show their maximum effects only in those situations in which children are from a single ethnic background. This may be a function of the extent to which parents are involved in the programs, the extent to which the social dynamics of the classroom are utilized by the model to develop learning environments, and the extent to which school personnel support either integrated or non-integrated situations. In any event, the integrated status of the classroom may effect the performance of the children in that classroom and this is a matter which must be considered when assessing the impact of any Sponsor.

3.4 GOAL 4: To Identify the Extent to Which the FT/NFT Contrasts Can be Attributed to the Unique Curriculum Inputs of Each Model.

The report of the analysis of this goal is included in Chapter VIII.

We have already drawn the distinction between a Sponsor's model and his program, and we have pointed out the likelihood that the effects of these two factors confounded in all the contrasts that we report here. We have also suggested that other situational factors may have confounding effects. We seek to describe, in Chapter IX, the extent to which the current analysis justifies attribution of effects to Sponsor models.

3.5 GOAL 5: To Describe Selected Characteristics of a Sample of Follow Through Teachers.

There are two goals in this section. The first is to produce a picture of some of the FT teachers from their responses to a mailed questionnaire. The items on this instrument covered demographic and training information, attitudes and values about the educational process, and some selected judgments about parent participation and the model with which the teachers are working. We expect that some of these variables will relate meaningfully to both the efficacy of model delivery within Sponsors, and to pupil outcomes. The first step in dealing with this issue is to determine the properties of the teachers and then to determine the distribution of these properties among Sponsors. The second step is to merge this data with pupil scores. To date, only the first step has been accomplished. In this report, therefore, we provide only a description of the teachers.

The second goal of this section is to attempt to identify some of the antecedents of teacher attitudes and their self-reported classroom behaviors in their training and demographic data. This will provide us with a fuller picture of the meaning of teacher attitudes and will give us a baseline for future estimates of Sponsors' effects on teachers. These results are reported in Monograph II.

3.6 GOAL 6: To Describe Selected Characteristics of a Sample of Follow Through Parents.

The purposes of this study are analogous to those of the teacher study and are reported in Monograph I. We wish ultimately to sort out the variance in pupil performance attributable to parental and home factors from that attributable to school and model factors. The parent interviews (conducted on a sample of FT parents by the National Opinion Research Council), covered a variety of demographic information, parental behaviors with the child at home, parental contacts with the school, and parental judgments of the school, the model, and the child's progress in school. It is clear that both parental approval of the programs, and parental participation in the educational process are goals of the FT program, and this data will help estimate the extent to which these goals have been and are being

reached. At the same time we wish to know if these factors vary from Sponsor to Sponsor in order to estimate the extent to which parental factors contribute to the model delivery and to pupil outcomes. Finally, it is important to discern some of the antecedents of parent attitudes and behaviors in the demographic data available in order to understand the nature of the problem that each Sponsor faces.

These data have not yet been merged with pupil data; this report describes selected parent properties, their distribution among Sponsors, and some interrelations among demographic and attitudinal data as they are distributed among Sponsors.

3.7 GOAL 7: To Describe Some of the Difficulties Which a Selected Set of Sponsors Have Encountered in Establishing and Administering Their Models in Selected Sites.

The purpose of this small study is to make the reader aware of the difficulties Sponsors have encountered when attempting to implement their models. The data for these descriptions are taken from a series of semi-structured interviews with central individuals located in the various Follow Through sites. They also include discussions held with representatives from these Sponsors. A short case study is provided in this report for each of a small number of Sponsors in a few sites. These data cannot be taken as a measure of the program, as previously defined, but simply as a demonstration of the importance of such measurement in assessing the impact of the model. Monograph III reports the results of this study.

This chapter has provided a short presentation of selected issues in the analyses of these data. It is designed to give the reader the information necessary to critically examine the analyses and the interpretations put to them. For those readers who wish an extended description of the linear model as used in this report, Monograph IV. discusses the theoretical aspects of the model, the assumptions under which it operates, and the particular procedures utilized in this study.

CHAPTER III

DESCRIPTION OF THE ANALYTIC SUBSET

1.0 INTRODUCTION

The primary analytic set of FT and NFT children investigated in detail for this report is drawn from the kindergarten, Cohort III portion of the data base. This subset is a group of children who have sufficient amounts of complete information to be useable in the analyses. Unlike a probabilistic sample, this subset has properties which may or may not be representative of either FT or NFT populations or of any Sponsor's portions of these groups.[1] The children, whose characteristics are described in detail in this section, entered school in Fall 1971 and were used to address Goal I in the analysis plan: to identify FT/NFT contrasts at the school, class, and child levels of aggregation. Other subsets drawn for special studies included in this report are described in conjunction with those analyses.

The data which pertain to and describe the kindergarten, Cohort III children were collected through a variety of sources. The children were tested twice during the 1971-1972 school year. They received a battery of achievement tests (Fall and Spring) and affective measures (Spring only). Other data were collected from the parents and teachers of these children. The parent measures (parent participation, parent's perception of school receptivity, and parents' satisfaction with their children's progress) and the teacher measures (teacher values, attitudes, and reported behaviors; teacher satisfaction; and teacher's perceived faithfulness to the Sponsor's approach) are described in more detail in conjunction with the reports of analyses of these data.

2.0 CHARACTERISTICS

The characteristics included in this discussion are selected for their usefulness in identifying the demographic and geographic

---

[1] Whereas it is understood that the groups of data analyzed are actually subsets of the data base, the term sample is also used to refer to this subset.

SPONSOR IDENTIFICATION CODE

Sponsor Code No.	Sponsors
01	Self-sponsored
02	Far West Laboratory
03	University of Arizona
04	George Peabody College
05	Bank Street College
06	University of Georgia
07	University of Oregon
08	University of Kansas
09	High/Scope Foundation
10	University of Florida
11	Educational Development Center
12	University of Pittsburgh
13	New York University
14	Southwest Educational Development Laboratory
15	Parent Implemented
16	Undecided
17	Hampton Institute
18	Northeastern Illinois State College
19	Georgia State University
20	Responsive Environments Corporation
21	Southern University
22	California Department of Education
23	University of North Dakota
24	Afram Associates, Inc.
25	(number not used)
26	University of California (Riverside)
27	Western Behavioral Sciences Institute

variables which may be contributing to essential differences between the FT and NFT groups. A psychometric measure, the WRAT, administered in the Fall is included in the discussion as an estimate of the entry level achievement of these kindergarten children.

The criterion for inclusion in the studies at the various levels of analysis is the availability of sufficient numbers of valid scores on children, their parents, and their teachers. Three subsets of data drawn for analysis are in some respects unique. In fact, different but overlapping portions of the data base constitute the various subsets at the school, class, and child levels of analysis. In each instance FT and NFT groups are separated for comparison.

At the school level of analysis, the data represent means for the children who were tested in each school, separated by grade level and FT/NFT where appropriate. These school means when averaged are not weighted by the number of students in the school. Over 75% of the schools are represented by between 16 and 124 children. Less than 8% are represented by between 5 and 10 children. No schools were included with less than 3 children with complete information. All sets of complete data (pupil, parent, and teacher) for the items of interest are included in the school level analysis, with the above exception.

At the class level of analysis the data reported pertain to class means, provided there are five complete data sets for that class. The averages of the class means are unweighted by the number of pupils. Approximately 45% of the classes are represented by 5 through 10 children and 20% are represented by 15 through 24 children. In both the school and class level studies, children representing all minorities are included.

At the child level of analysis the data reported pertain to the children themselves with the exception of the variable "percent White," which is a class characteristic. At this level of aggregation, the means reported are unit-weighted by the number of children entering the analysis. The primary difference between this sample and those at the class and school level is that only Black and White children are included in this analysis. The other minorities, such as Chicanos and

Indians, were highly concentrated in a few Sponsors' districts and consequently no statistical procedure could adequately adjust for this distribution.

2.1 Geographic Distribution

One important consideration in this study of characteristics of the FT/NFT analytic subsets used in this report is the identification of potential geographic bias. The two geographic variables discussed here are region and city size. Both of these variables are used in some form as covariates in the analysis.

Four regions are considered: Northeast, North Central, South, and West (see Appendix, Table AIII-1 for states comprising each of these regions). Four city size types are also considered: large cities (200,000 or more population); medium cities (50,000 to 199,999); small cities (10,000 to 49,999); and rural areas (less than 10,000). Figure III-1 presents the number of sites included in the three levels of analyses for each Sponsor for each region and city size combination. This figure displays the incomplete "sampling design":

- There are no rural sites in the Northeast or West;
- There are no small city sites in the North Central region. (This is not a sampling artifact; there are no FT sites in this cell); and
- There are no medium city sites in the South.

Whereas no Sponsor is represented by fewer than three sites, no region by city size site combination includes all Sponsors. It should also be noted that this is a maximum sampling representation; there are some sites in which an FT or NFT group is not included at all levels of analysis.

Further identification of the geographic differences between the FT and NFT groups for each Sponsor can be seen in Tables III-1 and III-2. These tables show the number of schools, classes, and children and the proportions which these numbers represent for each Sponsor's FT or NFT group.

S P O N S O R

	02	03	05	07	08	09	10	11	12	14
NORTHEAST										
Large City	1	1	2	1	2	1	2	1		1
Medium City			1	1	1			2		
Small City		1	2					1	1	
Rural Area										
NORTH CENTRAL										
Large City	1	2			1	1			1	
Medium City	1	1		2						
Small City										
Rural Area					1		1		1	
SOUTH										
Large City					1		1	1		
Medium City										
Small City		1				1				
Rural Area										1
WEST										
Large City						2				
Medium City	2									
Small City						1	1			1
Rural Area										

Figure III - 1 : NUMBER OF SITES FOR EACH SPONSOR WITHIN EACH REGION AND CITY SIZE

Table III - 1

The Number and Proportion of Schools/Classes/Children
for each Sponsor within each Region by FT/NFT

SPONSOR	LEVEL OF ANALYSIS	NORTHEAST				NORTH CENTRAL				SOUTH				WEST				TOTAL	
		FT		NFT		FT		NFT		FT		NFT		FT		NFT		FT	NFT
		N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	N
2	School	6	20.7	4	20.0	9	31.0	8	40.0					14	48.3	8	40.0	29	20
	Class	8	21.6	5	17.2	10	27.0	10	34.5					19	51.4	14	48.3	37	29
	Child	47	15.6	36	11.9	121	40.1	108	35.6					134	44.4	159	52.5	302	303
3	School	5	23.8	4	20.0	13	61.9	12	60.0	3	14.3	4	20.0					21	20
	Class	10	26.3	7	26.9	22	57.9	15	57.7	6	15.8	4	15.4					38	26
	Child	76	24.0	49	18.0	169	53.3	155	58.0	72	22.7	61	23.0					317	265
5	School	16	100.0	15	100.0													16	15
	Class	25	100.0	11	100.0													25	11
	Child	212	100.0	131	100.0													212	131
7	School	4	36.4	4	40.0	7	63.6	6	60.0									11	10
	Class	10	47.6	[illegible]	25.0	11	52.4	9	75.0									21	12
	Child	58	37.9	21	17.6	95	62.1	98	82.4									153	119
8	School	12	60.0	8	53.3	4	20.0	4	26.7	4	20.0	3	20.0					20	15
	Class	16	47.1	2	16.7	11	32.4	7	58.3	7	20.6	3	25.0					34	12
	Child	137	44.8	19	17.3	121	39.5	69	62.7	48	15.7	22	20.0					306	110
9	School	1	9.1	1	8.3	2	18.2	2	16.7	1	9.1	3	25.0	7	63.6	6	50.0	11	12
	Class	4	16.0			4	16.0	3	17.6	3	12.0	4	23.5	14	56.0	10	58.8	25	17
	Child	32	18.5	8	5.6	30	17.3	18	12.5	29	16.8	68	47.2	82	47.4	50	34.7	173	144
10	School	2	11.8			1	5.9	1	11.1	9	52.9	7	77.8	5	29.4	1	11.1	17	9
	Class	4	17.4			3	13.0	3	33.3	12	52.2	6	66.7	4	17.4			23	9
	Child	19	7.8			52	21.4	52	51.5	91	37.4	49	48.5	81	33.3			243	101
11	School	12	92.3	12	100.0					1	7.7							13	12
	Class	17	85.0	12	80.0					3	15.0	3	20.0					20	15
	Child	221	88.4	147	100.0					29	11.6							250	147
12	School	4	44.4	4	36.4	5	55.5	7	63.6									9	11
	Class	6	30.0	3	27.3	14	70.0	8	72.7									20	11
	Child	75	29.1	44	27.5	183	70.9	116	72.5									258	160
14	School	3	33.3	3	37.5					3	33.3	3	37.5	3	33.3	2	25.0	9	8
	Class	1	8.3							5	41.7	3	42.9	6	50.0	4	57.1	12	7
	Child	8	10.7	3	4.9					53	70.7	29	47.5	14	18.7	29	47.5	75	61
Total	School	65	41.7	55	41.7	41	26.3	40	30.3	21	13.5	20	15.1	29	18.6	17	12.9	156	132
	Class	101	39.6	43	28.9	75	29.4	55	36.9	36	14.1	23	15.4	43	16.9	28	18.8	255	149
	Child	885	38.7	458	29.7	771	33.7	616	40.0	322	14.1	229	14.9	311	13.6	238	15.4	2289	1541

Table III - 2

The Number and Proportion of Schools/Classes/Children for Each Sponsor Within Each Citysize by FT/NFT

SPONSOR	LEVEL OF ANALYSIS	LARGE CITY FT N	LARGE CITY FT %	LARGE CITY NFT N	LARGE CITY NFT %	MEDIUM CITY FT N	MEDIUM CITY FT %	MEDIUM CITY NFT N	MEDIUM CITY NFT %	SMALL CITY FT N	SMALL CITY FT %	SMALL CITY NFT N	SMALL CITY NFT %	RURAL AREA FT N	RURAL AREA FT %	RURAL AREA NFT N	RURAL AREA NFT %	TOTAL FT N	TOTAL NFT N
2	School	11	37.9	9	45.0	18	62.1	11	55.0									29	20
	Class	13	35.1	10	34.5	24	64.9	19	65.5									37	29
	Child	103	34.1	79	26.1	199	65.9	224	73.9									302	303
3	School	11	52.4	10	50.0	4	19.0	4	20.0	6	28.6	6	30.0					21	20
	Class	18	47.4	12	46.2	10	26.3	6	23.1	10	26.3	8	30.8					38	26
	Child	146	46.1	128	48.3	66	20.8	45	17.0	105	33.1	92	34.7					317	265
5	School	4	25.0	5	33.3	4	25.0	3	20.0	8	50.0	7	46.7					16	15
	Class	8	32.0	3	27.3	6	24.0			11	44.0	8	72.7					25	11
	Child	50	23.6	13	9.9	57	26.9	33	25.2	105	49.5	85	64.9					212	131
7	School			1	10.0	11	100.0	9	90.0									11	10
	Class	3	14.3			18	85.7	12	100.0									21	12
	Child			2	1.7	153	100.0	117	98.3									153	119
8	School	11	55.0	8	53.3	8	40.0	5	33.3					1	5.0	2	13.3	20	15
	Class	23	67.6	8	66.7	7	20.6	1	8.3					4	11.8	3	25.0	34	12
	Child	186	60.8	66	60.0	52	17.0	12	10.9					68	22.2	32	29.1	306	110
9	School	9	81.8	8	66.7					2	18.2	4	33.3					11	12
	Class	19	76.0	11	64.7					6	24.0	6	35.3					25	17
	Child	124	71.7	66	45.8					49	28.3	78	54.2					173	144
10	School	11	64.7	7	77.8					5	29.4	1	11.1	1	5.9	1	11.1	17	9
	Class	16	69.6	6	66.7					4	17.4			3	13.0	3	33.3	23	9
	Child	110	45.3	49	48.5					81	33.3			52	21.4	52	51.5	243	101
11	School	4	30.8	3	25.0	8	61.5	6	50.0	1	7.7	3	25.0					13	12
	Class	9	45.0	5	33.3	7	35.0	6	40.0	4	20.0	4	26.7					20	15
	Child	109	43.6	11	7.5	92	36.8	71	48.3	49	19.6	65	44.2					250	147
12	School	2	22.2	4	36.4					4	44.4	4	36.4	3	33.3	3	27.3	9	11
	Class	6	30.0	3	27.3					6	30.0	3	27.3	8	40.0	5	45.5	20	11
	Child	70	27.1	25	15.6					75	29.1	44	27.5	113	43.8	91	56.9	258	160
14	School	3	33.3	3	37.5					3	33.3	2	25.0	3	33.3	3	37.5	9	8
	Class	1	8.3							6	50.0	4	57.1	5	41.7	3	42.9	12	7
	Child	8	10.7	3	4.9					14	18.7	29	47.5	53	70.7	29	47.5	75	61
Total	School	66	42.3	58	43.9	53	34.0	38	28.8	29	18.6	27	20.5	8	5.1	9	6.8	156	132
	Class	116	45.5	58	38.9	72	28.2	44	29.5	47	18.4	33	22.1	20	7.8	14	9.4	255	149
	Child	906	39.6	442	28.7	619	27.0	502	32.6	478	20.9	393	25.5	286	12.5	204	13.2	2289	1541

From these tables the disproportional representation of the Northeast region and the large cities is evident. Tables III-1 and III-2 further show that at the various levels of analysis (school, class, and child), these two variables account for approximately 40% of the Sponsors' FT samples: the proportion located in the overall Northeast region ranges from 0.39 at child level, to 0.40 at class level, to 0.42 at school level; the proportion of the average Sponsor sample located in large cities ranges from 0.40 at the child level, to 0.42 at the school level, to 0.46 at the class level. The NFT samples have a much greater range in the average proportion at the various levels of analysis: Northeast region ranges from 0.29 at class level, to 0.30 at child level, to 0.42 at the school level of analysis; and large city size ranges from 0.29 at child level, to 0.39 at class level, to 0.44 at school level of analysis. At the class and child levels the average proportion of the NFT subsets in the North Central region exceeds that in the Northeast region. This is indicative of an overall FT/NFT disproportionality which varies with the samples at the different levels of analysis and may have effects on the results. The effects of this disproportionality are discussed more extensively in the results section.

In general, at the school level of analysis, there is a reasonable ratio between the FT and NFT samples on both of the geographic variables with no more than a 5.4% difference between the FT/NFT samples, averaged across Sponsors. Similarly, the average proportionality between the FT and NFT samples for each city size does not vary considerably at the different levels of analysis with the exception of large cities.

Whereas these average FT/NFT proportionalities across Sponsors indicate few outstanding differences, Tables III-1 and III-2 indicate considerable variation among Sponsors. A comparison of the FT and NFT groups at each level of analysis for each Sponsor on these two geographic variables follows, using Tables III-1 and III-2 as well as Figure III-1.

2.1.1 Sponsor 2 (Far West Laboratory)

There are no small cities or rural areas, nor any Southern sites represented in this Sponsor's subset. Otherwise, the geographic distribution pattern shows one site in a large city in the Northeast; two

sites in the North Central region, one in a large city and another in a medium-sized city and two sites in medium cities in the West. These are the only Western medium cities in the total subset. At the various levels of analysis there are no major changes in the distribution of the sample. There are some minor changes in the FT to NFT ratios but probably none sufficient to affect a systematic bias due to lack of an adequately representative comparison group for the FT groups.

2.1.2 Sponsor 3 (University of Arizona)

This Sponsor has no rural or Western sites in the subset. There are two sites in the Northeast, one large city and one small city. There are three sites in the North Central region consisting of two large cities and one medium city; and one small city in the South. The distribution patterns comparing FT and NFT proportions of the subset at each level of analysis indicate no changes in the site representation. Consequently, Sponsor 3 appears to have similar FT and NFT groups on the geographic dimension.

2.1.3 Sponsor 5 (Bank Street College)

This Sponsor's subset is totally located in the Northeast with two large cities, one medium city, and two small-city sites. The distribution pattern of the FT and NFT groups at the school level of analysis is fairly similar. At the class level, however, there are no NFT classes for comparison at the medium city site. Consequently, a large proportion of the NFT classes are located in the small city sites. At the child level, disproportionality between the FT and NFT subsets exists on the geographic variable, city size. In general, there is doubtful comparability among the subsets at the class and child levels of analysis.

2.1.4 Sponsor 7 (University of Oregon)

Sponsor 7 sites are almost exclusively in medium-sized cities; one in the Northeast and two in the North Central region. There is one large city site in the Northeast, a small proportion of the subset at all three levels. At the school and child levels, there are only NFT groups; at the class levels, there are only FT children. Consequently,

this large city FT to NFT ratio is highly variable and without a comparison within the Sponsor at any level. The medium city FT to NFT comparison is much more reasonable. Comparisons of FT to NFT groups by region show considerably more comparability.

2.1.5 Sponsor 8 (University of Kansas)

There are no Western or small city sites for this Sponsor. There is variation in the proportions of the FT/NFT sample at the various levels of analysis. However, the overall patterns tend to be consistent with excesses in similar directions at all levels. Note that the overall NFT sample is much smaller than the FT sample. At all levels over 50% of the FT/NFT groups are located in large cities: at two sites in the Northeast, and at one site in both the North Central and Southern regions. In addition, the greatest proportion of this Sponsor's FT sample is located in the Northeast at a total of three sites. A smaller proportion of the school level FT subset is located at the rural North Central site and the large city Southern site. At the class and child levels of analysis there is a marked disproportionability in the FT/NFT samples, particularly in the Northeast and North Central regions.

2.1.6 Sponsor 9 (High/Scope Foundation)

This Sponsor does not include any medium cities or rural areas, although all geographic regions are represented. There is one large city site in each of the Northeast and North Central regions, and two large city sites in the West (the only Western large cities in the analytic sample); there is one small city site in each of the Southern and Western regions. With the exception of the class level sample which has no NFT classrooms in the Northeast, there is a disproportion between the FT and NFT subsets for these geographic variables which remains consistent in all levels of analysis. There are consistently more FT than NFT groups from the large cities and fewer FT than NFT groups from small cities at all levels of analysis. The majority of Sponsor 9's subset is from those two large cities in the West, a unique occurrence for the analytic subset.

2.1.7 Sponsor 10 (University of Florida)

this Sponsor has no medium cities, and no NFT groups in the large cities in the Northeast or in small cities in the West except for one school included in the school level analysis in the latter instance. Similar to Sponsor 9, Sponsor 10 has sites in the four geographic regions. Two sites are located in large cities in the Northeast, one in a rural area in the North Central region, one in a large city in the South, and one in a small city in the West. Where an NFT comparison exists, there are relatively similar patterns in the FT to NFT ratios at the three levels of analysis. The distribution of the NFT sample is definitely unrepresentative; the majority of this group is in the large city in the South. Any FT to NFT comparison for Sponsor 10 probably warrants caution because of the uneven geographic distribution of the FT and NFT samples.

2.1.8 Sponsor 11 (Educational Development Center)

This Sponsor is primarily represented by four sites in the Northeast: one in a large city, two in medium-sized cities, and one in a small city. There is one additional site in a large city in the South. Consequently there are no North Central, Western, or rural sites. The ratio of FT to NFT groups is particularly disrupted by the lack of an NFT group in the large Southern city for the analyses at the school and child levels. In general, there are more FT than NFT groups in the large cities, approximately equal numbers in the medium cities and more NFT groups in the small cities. These ratios of disproportionality are evident at all levels of analysis.

2.1.9 Sponsor 12 (University of Pittsburgh)

Sponsor 12 is the only Sponsor in the subset with no large city sites in the Northeast. There are sites in one small city in the Northeast and in one large city and rural area in the North Central region. There are no Southern, Western, or medium city sites. The ratio of the FT group to the NFT group is fairly consistent at all levels of aggregation.

2.1.10 Sponsor 14 (Southwest Educational Development Laboratory)

This Sponsor is not represented by any sites in the North Central region or in medium cities. Like Sponsor 12, Sponsor 14 is only located in three sites in these analyses: one large city in the Northeast, one rural area in the South, and one small city in the West. The ratio between FT and NFT varies with the level of analysis. That is, there is no NFT at the class level for the large Northeastern city; at the child level there are more FT than NFT children in the rural area in the South, and the reverse relationship occurs in the small city in the West; and there is an approximately even distribution of children at the school level. With these varying FT/NFT proportions, it is doubtful that the subsets are similar at the various levels of analysis.

2.2 Demographic Descriptions

The Sponsor's FT and NFT subsets are described in this discussion using one psychometric variable (the special version of the WRAT administered in the Fall) and three demographic variables (percent White in the class or school, adjusted income of the child's family, and whether or not the child's mother has a high school education). These variables are covariates in the analyses. The overall means for the variables reported in Table III-3 shows that the FT and NFT subsets at the three levels of analysis are very similar. In general, the NFT's have a slight mean advantage at each level. There is also a tendency for a decrease in the level of the mean at increasing levels of analysis. This appears to be primarily attributable to differences between the subsets.

The Fall WRAT, as used in this description, is simply a pretest identifying entry level for the FT and NFT samples. This WRAT is a special version developed by USOE for administration to FT and NFT children. Table III-3 shows relatively little difference between the overall FT/NFT means on the Fall WRAT. These stable trends in actuality mask a considerable degree of variability within and among Sponsors. These differences will be discussed in more detail in the results section of this report.

Table III - 3

School/Class/Child Sample Size and Means on Four Covariates
For Each Sponsor by FT/NFT

SPONSOR	LEVEL OF ANALYSIS	FALL WRAT				PERCENT WHITE				ADJUSTED INCOME INDEX				MOTHER'S EDUCATION			
		FT		NFT		FT		NFT		FT		NFT		FT		NFT	
		N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$	N	$\overline{X}$
2	School	29	34.822	20	34.009	29	.320	21	.411	29	9.514	20	10.331	29	.515	20	.482
	Class	37	35.804	29	36.108	37	.302	29	.496	37	9.751	29	11.298	37	.498	29	.506
	Child	302	37.401	303	36.802	302	.366	303	.571	302	10.427	303	12.109	302	.590	303	.574
3	School	21	35.589	20	38.704	21	.535	20	.686	21	11.520	20	14.862	21	.527	20	.660
	Class	38	35.893	26	38.594	38	.479	26	.674	38	11.902	26	14.795	38	.563	26	.667
	Child	317	35.956	265	40.132	317	.485	265	.751	317	12.103	265	15.611	317	.618	265	.739
5	School	16	32.137	15	35.292	26	.461	22	.618	16	11.502	17	14.279	16	.468	17	.584
	Class	25	35.386	11	37.742	25	.587	11	.867	25	11.212	11	15.623	25	.477	11	.640
	Child	212	34.722	131	37.763	212	.671	131	.809	212	11.764	131	15.198	212	.524	131	.656
7	School	11	34.470	10	33.627	17	.136	14	.276	11	7.848	10	12.123	11	.429	10	.478
	Class	21	37.990	12	36.069	21	.190	12	.322	21	8.063	12	13.203	21	.418	12	.538
	Child	153	37.817	119	36.076	153	.208	119	.266	153	7.504	119	13.084	153	.438	119	.563
8	School	20	32.676	15	27.729	22	.082	16	.106	20	8.527	16	9.630	20	.409	16	.332
	Class	34	35.820	12	30.480	34	.084	12	.189	34	9.124	12	9.527	34	.442	12	.329
	Child	306	36.320	110	29.727	306	.151	110	.206	306	9.337	110	10.191	306	.461	110	.409
9	School	11	33.585	12	33.999	11	.189	13	.401	11	8.651	12	12.181	11	.500	12	.468
	Class	25	35.001	17	35.047	25	.136	17	.381	25	9.446	17	12.553	25	.514	17	.494
	Child	173	34.071	144	32.590	173	.138	144	.525	173	9.682	144	13.208	173	.601	144	.569
10	School	17	32.642	9	29.337	17	.352	9	.399	17	9.481	9	11.184	17	.390	9	.353
	Class	23	34.305	9	32.003	23	.365	9	.533	23	10.467	9	12.698	23	.407	9	.489
	Child	243	33.436	101	34.970	243	.535	101	.704	243	11.321	101	14.614	243	.469	101	.574
11	School	13	35.055	12	37.081	13	.553	12	.644	13	12.165	12	14.126	13	.538	12	.590
	Class	20	34.379	15	36.358	20	.359	15	.504	20	11.671	15	13.347	20	.512	15	.523
	Child	250	34.496	147	37.436	250	.422	147	.770	250	12.088	147	13.912	250	.528	147	.612
12	School	9	36.182	11	36.261	9	.779	11	.717	9	10.821	11	14.167	9	.653	11	.650
	Class	20	37.933	11	36.524	20	.701	11	.770	20	11.498	11	14.658	20	.671	11	.737
	Child	258	38.578	160	37.831	258	.730	160	.859	258	11.682	160	14.988	258	.713	160	.775
14	School	9	26.991	8	29.599	9	.226	8	.374	9	8.252	8	10.059	9	.273	8	.344
	Class	12	31.308	7	29.436	12	.281	7	.470	12	9.803	7	10.982	12	.390	7	.392
	Child	75	31.760	61	29.902	75	.385	61	.437	75	10.347	61	11.705	75	.493	61	.426
Overall Means	School	156	33.671	132	34.004	174	.359	146	.472	156	9.903	135	12.404	156	.473	135	.503
	Class	255	35.588	149	35.579	255	.342	149	.525	255	10.317	149	12.306	255	.495	149	.543
	Child	2289	35.794	1541	36.276	2289	.420	1541	.621	2289	10.801	1541	13.634	2289	.552	1541	.615
Overall Standard Deviation	School		5.325		6.720		.394		.414		2.723		3.483		.180		.225
	Class		5.997		6.528		.397		.397		3.101		3.645		.200		.222
	Child		13.496		14.053		.425		.390		6.337		6.743		.497		.487

A second demographic variable considered as a descriptor is the racial composition of the class or school. This variable indicates the proportion of White students in the class or school. At the child level, the score for an individual student is the proportion of Whites in his entire class, so all students in a classroom receive the same score for this variable. Overall, the NFT group has more White children per class or school than the FT group. This difference exists at all levels of analysis and within each Sponsor (with the exception of the school level analysis for Sponsor 12). The FT to NFT overall difference ranges from 0.11 at the school level to 0.20 at the child level. A further investigation showed that some Sponsors work with predominantly "White" and "non-White" classes and schools, with few having means close to the overall mean values. Such a result may be a function of each Sponsor's differential appeal to various types of communities.

The third demographic variable used to describe these subsets is the adjusted income index. This variable takes into account the income of the family, the number of people in the household and whether the family lives in a rural area. The range of the variable is zero to 25, the upper values representing a relatively higher ratio of the family's adjusted income to a subsistence level. Overall, there is a great deal of variability among Sponsor means at the different levels of analysis despite the relatively stable FT to NFT mean ratios. In fact, FT groups have lower incomes than NFT groups, at all levels of analysis, as would be expected by the eligibility guidelines for the FT program. There are "rich" and "poor" Sponsors. Some Sponsors have widely differing FT and NFT groups; others have well-matched groups. This variability will be discussed in more depth as it affects each Sponsor individually.

Mother's education, used as a dichotomous variable identifying whether the child's mother has completed a high school education, is a fourth demographic variable used to describe the FT to NFT contrasts. Average scores reported here represent the proportion of students at each level of analysis whose mothers have completed high school or have a more advanced education. Overall, the average FT score for each level

of analysis is lower than the average NFT score, although there is considerable variability of that ordering within Sponsors. This Sponsor FT to NFT variability is discussed in detail in the results which follow in Chapter VII-4.1 to 4.10.

3.0 OUTCOME VARIABLES

The outcome measures analyzed in the various kindergarten studies include academic achievement tests administered in Fall 1971 and Spring 1972, measures of motivational orientation taken from the Spring 1972 kindergarten test battery, and a measure of absence. Internal consistency reliabilities are given for each of these measures (except absence) using Hoyt's analysis of variance estimate of reliability (Hoyt, 1941).

3.1 Academic Achievement

- Wide Range Achievement Test (WRAT). This test is an individually administered measure of letter and number recognition, word reading, spelling, and oral and written arithmetic problems. Whereas national norms are available, the spelling, reading, and arithmetic subtests were abbreviated for administration in the Follow Through evaluation. Thus, the published norms are not applicable to this modified version. The scores analyzed here are total raw scores with a range from 0 up to a maximum score of 84. On a sample of 4,769 kindergarten children, a reliability of .92 was obtained. The Fall WRAT served as the pretest covariate for all outcome variables except the Peabody Picture Vocabulary Test.

- Peabody Picture Vocabulary Test (PPVT). This is an individually administered measure of picture recognition vocabulary. The child is required to point to the correct picture corresponding to the word spoken by the examiner. The test was administered according to standard procedures; however, a number of pictures were modified to reflect Black rather than White ethnic characteristics. Thus, the test scores are not directly parallel to those derived from the standard version. Total scores equal the number of correct responses.

  The Hoyt reliability estimate was .89 in the Spring 1972. The same test administered in Fall 1971 served as the pretest covariate for the Spring PPVT.

- Metropolitan Achievement Tests (MAT). Three subtests from this multiple choice, group-administered test were analyzed as dependent variables: Reading, Listening for Sounds, and Arithmetic. The first is a test of letter and word reading. The second is a test of recognition of initial, medial, and final word sounds, representing important reading readiness skills. The Arithmetic subtest is a measure of basic math concepts, vocabulary, and addition and subtraction skills. National norms are available for these subtests administered in the Spring 1972 testing; however, the results reported here are in terms of raw scores.

  The Hoyt reliability estimates obtained for these subtests were .83 for Reading, .86 for Listening, and .87 for Arithmetic.

3.2 Motivational Orientation

- Gumpgookies. This test, developed by Adkins and Ballif, measures the child's motivation to achieve in school. The hypothetical constructs underlying this measure are theorized to be a dynamic interaction of learned responses unrelated to intellectual ability, including: the child's knowing and performing activities directed toward achievement; enjoyment of the school situation; self-evaluation; self-confidence in physical activities; and the child's purposive behavior toward accomplishing future goals. As an individually administered picture test, the child is required to point to one of two semi-projective "Gumpgookie" figures which presumably reflects the orientation of the child either toward or away from one of the five dimensions outlined above. The Gumpgookies are vague figures with the outline of a head, arms, and legs, resembling "Casper" the ghost.

  Total raw test scores were analyzed for these studies, the maximum score possible being 60. The Hoyt reliability estimate obtained in the same sample referred to above was .88.

- Locus of Control (Locus-positive and Locus-negative). This individually administered picture test, developed by Shipman, reflects the child's perception of the extent to which he, or others in his environment, are responsible for the events that happen to him -- a hypothetical dimension first explored by Rotter and later substantiated by a wide body of research. The events characterized in this test are restricted to the academic and social situations in the child's school

life. Two scores are derived from the test Locus-positive (11 items), reflecting the child's perceived responsibility for good events; and Locus-negative (9 items), reflecting the child's responsibility for unfavorable happenings in his school situation. Research on this test suggests that different intra-personal dynamics underlie these two scores; thus the total Locus of Control score is not separately analyzed here. Hoyt reliability estimates obtained from the sample described above were .43 for Locus-positive and .22 for Locus-negative. These low indices of test internal consistency reflect not only the short test length of these subscales, but also the imperfect state of the art in measurement of the affective domain in children. One effect on the analyses of these low reliabilities is to restrict the possible proportion of total test variance that may be accounted for by the predictor model, and thus reduce the possibility of detecting significant, though imperfectly measured, motivational relationships in the test performance patterns. Thus any effects which actually are obtained on these measures may be considered underestimates of the true relationships underlying the data.

3.3 Absence

- The number of days a child missed school throughout the kindergarten year was analyzed as an indirect measure of a general capacity for a Sponsor to attract pupils consistently. While children who enjoy school are more apt to attend, this interpretation cannot be fully accepted without qualifications concerning other reasonable determinants of absence such as sickness, parental factors, or environmental phenomena operating within the community.

CHAPTER IV

COVARIATES

1.0 INTRODUCTION

Every assessment of the effects of an experimental program or treatment is confronted with the problem of confounding: extraneous factors (i.e., factors not part of the treatment conditions) may often be related both to the outcomes under study and also to the treatment conditions themselves. When confounding occurs the effects of the treatment are mixed with the effects of the extraneous factors and any analysis which ascribes all changes to the treatment may seriously underestimate or overestimate the true treatment effects.

In a controlled experimental situation two basic procedures minimize the danger of confounding:

- rigorous control over the administration of the experimental treatment conditions, ensuring that all subjects in a specified group receive the same treatment in the same manner; and
- random assignment of subjects to experimental conditions, so that the effects of extraneous factors associated with individuals affect treatment and control groups equally, within statistically determinable limits.

When an experiment takes place in a natural setting the experimenter generally cannot escape the effects of confounding in these two standard ways. FT, in particular, is a quasi-experiment being performed under real-life conditions. Subjects could not be randomly assigned to treatment conditions, and rigorous control over the administration of the treatments could not be maintained. We must, therefore, deal explicitly in this report with the problem of confounding.

Two general categories of confounding factors potentially affect the results of the FT experiment: aspects of implementation and Sponsor delivery, and nonrepresentative sampling. These arise from the unavoidable problem of controlling treatments and randomizing subjects.

Problems of implementation arise as Sponsors attempt to apply their educational models to realities of the lives of children in public schools. Each Sponsor has tried to implement his program in a variety of sites. Local history and circumstances, extraneous to Sponsor intentions, have caused different sites to respond very differently, even to the same model

presented in the same way. In some sites, Sponsors have tailored their approach to their perceptions of local needs and constraints. Some Sponsors' models, indeed, contain explicit provisions for flexibility and adaptability. Site-to-site variations in a given Sponsor's model may well make excellent educational sense, but they complicate the evaluator's task by introducing effects that he must regard as extraneous. Sponsors vary, furthermore, in their ability to implement their models faithfully, even in equally cooperative sites. The effects of these variations are not altogether extraneous: we may judge a Sponsor, in part, on his ability to translate his theories into action.

Even if Sponsors' models were uniformly and faithfully implemented, and even if the implementation processes had not activated any site-specific confounding influences, we should still wish to ensure that non-Follow Through subjects be reasonably comparable to Follow Through subjects on relevant characteristics--both within and across Sponsors--to avoid confounding artifacts of selection with the FT/NFT and Sponsor comparisons we seek to make. By relevant characteristics we mean characteristics correlated with the outcome measures of interest.

Follow Through and comparison populations do differ in a number of important ways, both within and between Sponsors. The description of the sample, presented below, documents some dimensions of this variation. At this point it suffices to note this situation and to point out that we have attempted to adjust statistically for this non-comparability by means of the analysis of covariance. A technical discussion of this procedure, together with some problems and pitfalls inherent in its use under the present circumstances, appears in Monograph IV. We now turn to a discussion of the factors which define non-comparability across groups and which we have therefore employed as covariates in the present set of analyses.

2.0 WHICH COVARIATES?

Six basic categories of extraneous factors may have effects which are potentially confounded with FT and Sponsor effects. Each category is represented to some extent in the set of covariates that we have employed in the analyses which follow:

- pupil characteristics, such as individual pupils' entry scores on achievement tests, ethnicity, etc.
- parent/family characteristics, such as mother's education, SES of the family, the degree to which parents participate in school activities, length of time residing at current address, etc.
- teacher characteristics, such as years of education, years of teaching experience, ethnicity, etc.
- class characteristics, such as aggregated factors of the ethnic mix of the pupils in the class, the mean entry score on achievement tests, the mean achievement motivation score, etc.
- school characteristics, which may be further classified into two categories: (1) aggregated factors like those indicated for classroom characteristics, and (2) global characteristics which are defined independently of the characteristics of individual persons, such as pupil teacher ratio; the presence or degree of use of specialized professional staff such as psychologists, speech therapists, reading specialists the presence of a subsidized such program or bussing for the purpose of achieving racial balance; etc. Unfortunately there is currently no data on global characteristics of schools available, although efforts to collect such data are presently being mounted; we shall use the results in future analyses.
- environmental characteristics, such as region of the country and the size of the city in which a site is located.

These six categories of factors define a large universe of potential covariates for the analyses that follow. By no means were all of these variables measured or even measurable in FT. We selected our final list of 18 variables by applying five sequential constraints in approximately the following order: (1) the relevance of the variable as suggested by previous research, (2) our own thinking about the theoretical and methodological problems involved in the analyses, (3) the availability of the variable in the data base, (4) the requirement that a variable to be included correlate at a reasonable level with some outcome measure in the analyses, and (5) that the variable behave homogeneously across the various treatment groups; that is, that it not interact significantly with Sponsor and treatment variables.[1]

[1] Our samples are too small to permit us to introduce appropriate interaction terms in the model to adjust for covariate heterogeneity.

2.1 Levels of Analysis and the Meaning of the Covariates

The studies which constitute the assessment of FT have been conducted at three distinct levels of analysis: (1) at the child level of analysis, (2) at the class level of analysis, and (3) at the school level of analysis. Note that when schools are indicated as the unit of analysis we do not literally mean the entire school. Rather, we refer only to the tested children within the school.

While interpreting our results, the reader must keep in mind the appropriate frame of reference: similar-looking studies at different levels of analysis do not address the same questions. At the child level of analysis, a study examines the impact of FT on the performance of individual children, taking into account their own sets of personal, unique characteristics. At the class level of analysis, on the other hand, the performance of individual children is no longer a point of consideration. When we operate mathematically on the characteristics of the children in a class by computing a mean or a proportion, the resulting variable is no longer a characteristic of any individual child but rather of the class as an entity in its own right. Class studies examine the behavior of classes in terms of class characteristics, not in terms of the individual characteristics of the members of the class. This distinction is important to keep in mind, for relationships and processes do not necessarily exist at macro, or aggregated, levels of analysis that hold at the individual level of analysis. Relationships observed at a macro level of analysis such as the class or school using aggregated variables may be weaker, stronger or even the reverse of the analogous relationships among individuals. It is therefore not safe to assume, except under an extremely restricted set of conditions, that the school or class level analyses simply replicate the child level analyses.

As we have already suggested, moreover, variables formed through mathematical operations at one level of aggregation may be used unchanged but with different meanings at other levels of analysis.

Consider, for example, some possible uses and conceptual meanings of the Fall WRAT aggregated to a class mean. At the class level of analysis the class mean on the Fall WRAT is a characteristic of the class *per se*

and indicates the academic level of the class at entry in the Fall. As such the class mean serves as the class pretest score to adjust for differences in academic starting levels among classes when we study the behavior of classes. This use and meaning of the class mean is conceptually parallel to the use of individual pupils' Fall WRAT scores to adjust for initial differences in academic starting levels among children when we study the behavior of children.

The class mean on the Fall WRAT may be used for an entirely different purpose, however, with a different meaning at the child level of analysis (i.e., when studying the behavior of children). When applied at the child level of analysis, the mean of the class in which each child is located may serve as a contextual variable: an indicator of the immediate environment or milieu in which the child receives his schooling. Since human beings act upon and are affected by their environments, we expect that any specific classroom environment will affect the children within the class differently depending on their personal characteristics; different classroom environments may have different effects on children in general. The mean academic entry level of a class represents one such aspect of the classroom environment.

Keeping these distinctions in mind, we now turn to a discussion of the meaning of each of the covariates employed in the studies that follow.

3.0 THE COVARIATES USED IN OUR ANALYSES

Table IV-1 presents an overview of the 18 variables employed as covariates at the various levels of analysis in the studies that follow. We shall address each of these variables in turn, in terms of its meaning at each of the levels of analysis at which it is employed.

(1) Fall WRAT

At the child level of analysis the child's score on the Fall administration of the WRAT provides an indicator of his academic starting position upon entry into the FT program. At the class and school levels of analysis,

Table IV-1. Overview of Variables Used as Covariates in Studies at Three Levels of Analysis

COVARIATE	PUPIL LEVEL OF ANALYSIS	CLASS LEVEL OF ANALYSIS	SCHOOL LEVEL OF ANALYSIS
1. FALL WRAT	Child's score	Mean score of (tested) children in the class	Mean score of (tested) children in the school
2. PRESCHOOL EXPERIENCE	Binary variable (any type of pre-school experience vs. none)	N/A	N/A
3. MOTHER'S EDUCATION	Binary variable (high school or more vs. less than high school)	% of mothers of tested children in the class with a high school education or more	% of mothers of tested children in the school with a high school education or more
4. ADJUSTED INCOME INDEX	Parent's score on Adjusted Income Index	Mean score for tested children in the class	Mean score for tested children in the school
5. PARENTS' PERCEPTION OF SCHOOL'S RECEPTIVITY TO PARENT INVOLVEMENT	Parent's score on Receptivity Index	Mean score of parents of tested children in the class	Mean score of parents of tested children in the school
6. PARENT PARTICIPATION IN SCHOOL ACTIVITIES	Parent's score on Participation Index	Mean score of parents of tested children in the class	Mean score of parents of tested children in the school
7. YEARS AT CURRENT ADDRESS	Number of years family has lived at current address	Mean years at current address for tested children in the class	Mean years at current address for tested children in the school
8. TEACHER'S EDUCATION	Binary variable (bachelor's degree or higher vs. less than a bachelor's degree)	Binary variable (bachelor's degree or higher vs. less than a bachelor's degree)	N/A
9. TEACHER'S TEACHING EXPERIENCE	Number of years of teaching experience	Number of years of teaching experience	N/A
10. ETHNICITY OF TEACHER	N/A	Binary variable (Black vs. non-Black)	N/A
11. FALL WRAT (STANDARD DEVIATION)	N/A	Standard deviation of Fall WRAT scores of tested children in the class	N/A
12. INTEGRATION	% of tested children in the class who are white	% of tested children in the class who are white	% of tested children in the school who are Black
13. % MINORITY	N/A	N/A	% of tested children in the school who are of a minority group
14. CITY SIZE	4 categories (≥ 200,000; 50,000 ≤ 200,000; 10,000 ≤ 50,000; < 10,000)	4 categories (≥ 200,000; 50,000 ≤ 200,000; 10,000 ≤ 50,000; < 10,000)	N/A
15. METROPOLITAN AREA	N/A	N/A	Binary variable indicating whether or not school is located in an SMSA
16. MIDDLE-SIZED CITY	N/A	N/A	Binary variable indicating whether or not school is located in a middle-sized city
17. WESTERN REGION	N/A	N/A	Binary variable indicating whether or not school is located in the Western Region
18. SOUTHERN REGION	N/A	N/A	Binary variable indicating whether or not school is located in the Southern Region

the mean score on the WRAT is computed for the appropriate children. The class and school means carry parallel meaning: it is the average starting point for a group of children.

(2) Preschool Experience

Preschool experience, used only at the child level of analysis, is a binary variable indicating whether or not a child has had any preschool training. This is used as a covariate on the assumption that previous experience in school may be positively correlated with performance in the FT program, thus making children with preschool experience appear more responsive to FT, or giving a spurious advantage to NFT or FT groups that have a high proportion of children with such experience.

(3) Mother's Education

At the child level of analysis, mother's education is coded as a binary variable indicating whether a child's mother has a high school diploma. We assumed that mothers with more education are more likely than mothers with less education to engage in interactions with their children that are conducive to the development of higher academic aptitudes, attitudes, and performance, as well as interactions that lead to a more positive affect on the part of the child, not only toward himself but toward school as well.

At the class and school levels of analysis, the variable is defined as the proportion of mothers with a high school education or more. At an aggregated level, a high proportion of mothers with advanced education extends beyond the implication that a high proportion of children in the class or school are exposed to the parent-child interactions indicated above. We reasoned that mothers with higher education are more likely to get actively involved in school affairs, visit teachers more often, have different expectations for the school, and relate to staff more often and in a qualitatively different way, than mothers with lower education. Indirectly then, the proportion of mothers with more education may well affect what goes on in the class or school in terms of the educational process, thereby affecting the performance of the class or school taken as a unit.

(4) Adjusted Income Index

This variable is operationally defined by NORC's Poverty Range Index, which is the ratio of the annual income of a family to a basic subsistence level computed by taking into consideration the number of persons in the household and its location in an urbanized or rural area. This index has a possible range of scores from 0 to 25.

At the child level of analysis, a person's score on the income index is a measure of his economic status. In terms of the present analysis this is a proxy for the processes, values, etc., associated with an economic status which affect the individual's affective development and academic aptitudes and attitudes.

At the class or school level of analysis, the mean score on the income index is an indicator of the climate of the class or school in much the same way as the proportion of mothers with higher education previously discussed. That is, different mean scores of various classes or schools on the income index probably indicate differences in climates (both internally as regarding the children directly, and externally as regarding the behavior or parents in relation to the teacher and school) that affect the educational process within the class or school in many ways.

(5) Parents' Perception of the School's Receptivity to Parent Involvement in School Activities

This variable is the same variable used in the parent studies; its specific operational definition is presented in Monograph I. At the child level of analysis, a given parent may perceive school receptivity to parent involvement quite idiosyncratically. Whether or not the parent's assessment of the situation is accurate, it may influence the way in which the parent relates to the school as well as his affect toward the school, which may be transmitted to the child. This, in turn, may influence the child's attitudes toward school and ultimately his performance.

At the aggregated levels, parents' mean score on the school receptivity measure probably provides a fairly accurate representation of the openness of the school to parental involvement. A high degree of openness to parental involvement may be indicative of friendly and cooperative parent-school relations, responsiveness to environmental pressures on the part of

the schools, etc. These factors may have many direct and indirect ramifications for the educational process within the class or school. Sponsors, furthermore, place varying degrees of emphasis on parent-school relations as a desired outcome.

(6) Parent Participation in School Activities

This variable is the same as the variable used in the parent studies and its precise operational definition is presented in Monograph I. At the child level of analysis, the extent to which a child's parents participate in school activities in an indicator of the parents' interest in the school and in the child's education. It also measures indirectly the availability of the resources that such participation requires. Both of these factors may have an impact on the child's affect toward school and performance in it.

At the class and school levels of analysis the mean level of parent participation in school affairs measures the climate of interest on the part of parents that may have diverse effects on both school and children; these may in turn affect the performance of the children in the school. and thereby the performance of classes or of the school in aggregate terms.

(7) Years at Current Address

At the child level of analysis the number of years a child has lived at his current address is an indicator of the geographic stability of his family. Conceptually this implies a hypothesis to the effect that there are processes which develop concomitantly with geographic stability which contribute in some way to the child's academic and affective development.

At the class or school level of analysis, the mean number of years that pupils have resided at their current residence is an indicator of the turnover of the student body. The degree of turnover of the student body probably affects several other characteristics of the school, such as the development of coherent and recognized norms among the pupils, the types and depths of friendships pupils can develop, the interaction patterns of school and parents, etc. All of these may influence indirectly the academic and affective development of the pupils in the school.

(8) Teacher's Education

Teacher's education is represented by a binary variable indicating whether a teacher has less than a bachelor's degree, or a bachelor's degree or higher. In the child and class level analyses, this variable is an indicator of the qualifications of the teacher.

(9) Teacher's Years of Teaching Experience

This variable is operationally defined as the number of years of teaching experience a teacher has. In the child and class level analyses the experience of the teacher is another indicator of qualifications.

(10) Ethnicity of the Teacher

Used only in the class level analyses, the ethnicity of the teacher is represented by a binary variable (Black versus non-Black). We assumed that teachers more similar to their pupils in cultural and ethnic background relate to their pupils better than teachers who are dissimilar.

(11) Standard Deviation of the Fall WRAT

Used only at the class level of analysis, the class standard deviation on the Fall WRAT provides an indicator of how homogeneous the academic entry level of the class was. We assume that classes composed of children with widely divergent academic entry levels behave differently from classes composed of children very similar in entry level. At the least, homogeneous classes present somewhat different tasks for the teacher from those presented by heterogeneous classes.

(12) Integration

This variable is operationally defined at the class level of aggregation as the precent of the children who are White, and at the school level as the percent of children who are Black. Percent White is used in the child level analyses as a contextual variable and at the class and school levels as a characteristic of the class (school) basically parallel in meaning to the ethnicity of an individual child at the pupil level of analysis.

The primary rationale for using ethnic group membership as a covariate is the assumption that the various racial/ethnic groups have differing values, norms, group structures and processes, interpersonal interaction patterns, and in general, live different life styles. At the individual level of

analysis, the factors associated with the racial/ethnic group of which the person is a member do, in fact, have a direct impact upon that individual personally through the socialization process. In the case of the FT assessment, the impact of these factors may relate to the types of academic factors measured in the FT instruments.

At the school level of analysis, the proportion of persons in the school having various racial/ethnic group memberships has implications beyond the simple fact that some given number of children are being personally and directly affected by their racial/ethnic group membership outside of school. A high proportion of a given racial/ethnic group in school may well establish a prevailing climate within the school which is reflective of the racial/ethnic group's values, norms, etc. This climate will have effects on all persons in the school, above and beyond their own racial/ethnic group membership.

The racial/ethnic composition of the school may also affect the work structure, policies, and organization of the school as well as the values, attitudes, and behavior of the school's staff and administration. Generally, the racial/ethnic composition of a school's student body also implies that the external environment of the school, as constituted by parents, has a comparable composition. This fact may ramify in terms of the manner in which parents and school relate to each other, the intensity and tenor of parent-school interactions, degree of community support of the school, etc.

(13) Percent of Children in the School Who are Members of Minority Groups

Used only at the school level of analysis, the percent of the children in a school who are members of a minority group has the same general implications as discussed in terms of percent Black in (12) above, with the operational definition broadening the scope of ethnic group membership to include all minority groups.

(14) Size of City

The size of the city in which a school is located is defined in four population categories: (1) under 10,000; (2) 10,000 to 49,999; (3) 50,000 to 199,999; and (4) 200,000 or more. Used as an indicator of

the broad environment in the child and class level analyses, size of city is a proxy variable carrying a diverse set of information relevant to the different norms, values, and processes existing in cities of different sizes.

(15) Metropolitan Area

This is a binary variable used only in the school level analyses, which indicates whether a school is located within or outside of a Standard Metropolitan Statistical Area. It is used in conjunction with the variables, "middle-sized city," "western region," and "southern region."

(16) Middle-sized City

This is also a binary variable used only in the school level analyses, indicating whether a school is located in a middle-sized city or not. These two variables replace, in the school level analyses, the size-of-city city variable used at the child and class levels of analysis, which proved to be hetergeneous (interactive) at the school level. However, basically the same information is carried in either operational definition.

(17) Western Region, and

(18) Southern Region

These are two binary variables used only at the school level of analysis, which indicate whether or not a school is located in the southern region or western region of the country (Northeast and North Central were not used because of interactions). These variables are indicators of the broad environment in which a school is located, as are the size-of-city variables discussed above. In this case the variables carry information related to the differences, on a number of factors, to be found among the different regions of the country.

CHAPTER V

METHODOLOGICAL PROBLEMS IN THE ANALYSIS OF THE FOLLOW THROUGH DATA

1.0 THE CAVEATS

The standard (classical) statistical methods of drawing inferences from data are variations on the theme of deciding whether some independent variable has a significant effect on the (set of) outcome measure(s) with a statistical distribution to help decide whether the effect actually occurred. If, for example, both the cause(s) and effect(s) are measured on a nominal scale, one would employ an appropriate version of the χ^2 test for deciding whether or not the observed differences are statistically significant. If both the cause(s) and effect(s) are continuously measurable and reasonably normally distributed, one usually employs a suitable F test for drawing conclusions about the propriety of postulated cause and effect relationships. A whole array of statistical procedures between χ^2 and F are available for testing the cause-effect relations at levels of measurement between nominal scales and normal distributions. All of these procedures have a common base: the assumption that the data are obtained from a well designed experiment. Typically, social science investigations are seldom experiments and Follow Through is not an exception to this rule.

Briefly, FT data suffers from five maladies: imbalance, missing data, fallibility of measures, non-homogeneity of responses, and non-probabilistic sampling. Consequently, even though statistical procedures have been used to guard against unwarranted conclusions, the results are presented (and should be interpreted) with a minimum use of statistical jargon. Before proceeding to report our results, however, a few comments are due on the technique used for data analysis, and the problems with the FT data.

1.1 The Covariance Technique

The goal of the standard analysis of covariance is to adjust for various kinds of potentially confounding group differences. Where treatment and comparison groups differ in size, and/or where the assumed uniformities do not in fact obtain, the results of the analyses

reflect not only the 'treatment effects' of interest but also varying amounts of extraneous "effects of unmet assumptions."

The FT data, needless to say, fit imperfectly into the ideal structures that standard analytic procedures assume, even though they were originally designed to satisfy those structures. The inadequacies in the FT data are listed below.

1.2 Imbalance

The FT data are unbalanced: that is, the Sponsors serve varying numbers of children, parents, classrooms, schools, and school districts; FT and NFT populations also vary in size, both within and across Sponsors. The observed frequencies (i.e., those participants about whom the records are available and complete) are neither equal nor proportional, and hence the analysis design is not orthogonal. Thus, when testing hypotheses or estimating the variations in outcome measures explained by changes in the predictor set, the order of testing (or estimating) is very important--the variables introduced earlier are given more than their share of credit for explaining the outcome measure variations.

1.3 Missing and Incomplete Data

The data sets are incomplete in a number of respects. An individual's record may include some scores but lack others. Children and projects have joined the experiment late or left it early. We have been particularly hampered by the unavailability of pretest scores on many outcome measures.

Some of the recorded data had to be discarded (and was thus "missing" for analysis) due to errors in the encoding process. Some children, for example, are recorded as female in one year and male in the next. Since the original records could not be reviewed, these types of observations were dropped--our only other choice being "random assignment" of sexual and other demographic characteristics.

1.4 Fallibility of Measures

By this we mean (1) that what was measured does not necessarily correspond to what was intended to be measured; and (2) that measurements are not sufficiently accurate to give the same results twice. Both of these problems exist to a great extent in the data.

1.5 Non-Homogeneity

The data are non-homogeneous, reflecting the persistent variety of the real situations in which FT models have been implemented. Analysis of covariance assumes that the variables in the analysis have approximately the same variability in one Sponsor's FT or NFT group as any other group defined by the design, and that the correlations among those variables also hold constant (except for measurement error) from one design group to another. Unfortunately, this doesn't hold very dependably in the FT data base.

There is also another aspect to the non-homogeneity, and this had to do with the behavior of the covariate set. Ideally, the covariate set should remove all, and only, the effects of initial differences between those chosen to receive Follow Through and those not so chosen. In practice, we found that some of the covariates have interactions with the treatments, i.e., the covariates provide different levels of adjustments depending on the level of treatments--and thus some of the covariates did not behave as statistical equalizers.

1.6 Non-Probabilistic Sampling

The data are non-probabilistic, since projects, classrooms, teachers, and children were selected judgementally for both Follow Through and non-Follow Through. This circumstance makes covariance adjustment both indispensable and very difficult.

The non-probabilistic sampling also affects generalizability of the findings. Intuitively, one may consider the population of potential participants to be divisible into a number of mutually exclusive groups. Random (probabilistic) sampling allows for a representation of each group; selective

sampling tends to choose certain groups, ignoring others. A selective or judgemental sampling procedure may not allow a fair representation of a variety of experiences and life-styles of potential participants. (Technically speaking, the predictor or causal variables tend to cluster together, thus reducing the variance in the denominator of F tests.) It would appear that under this scheme the standard F ratios are inflated and tend to show "significant" results more often than warranted. For this and similar reasons, we have chosen to answer each research question by applying more than one technique; by not relying heavily on probabilistic statements of significance; and by employing foreign-to-educational-research terminology such as "signal" and "noise." The choice and discussion of various techniques for delineating the FT effects now follows.

2.0 METHODS SELECTED FOR THE CURRENT ANALYSIS

The choice of analytic techniques was dictated by (1) the research questions of interest; (2) the sampling techniques employed for choosing the FT/NFT participants (children, classes, schools, parents, teachers); and (3) the state of the FT data recorded on the tapes received by Abt Associates data analysis staff.

In the absence of a single perfectly trustworthy model, it behooves us to approach each research question, wherever possible, from a number of analytical angles, balancing off the complementary advantages and drawbacks of parallel analyses in a cross-validation strategy. This report reflects the beginnings of such a strategy; later reports will take advantage of expanded data availability over time and across Cohort boundaries to introduce new, parallel analyses to corroborate or refine the results we present here.

Not all analytic procedures that measure experimental/control contrasts merit a place in our scheme. For example, a simple t test of an overall FT mean against the analogous NFT mean, while appealing in its straightforward simplicity, would not reflect Follow Through's vital sponsorship structure and is, therefore, not reported. A somewhat more complex procedure, the analysis of variance (ANOVA), captures sponsorship but assumes that the FT and NFT groups to be contrasted were initially equivalent before Follow Through intervened. In view of the extremely judgemental way in which non-Follow Through was selected, and in view of the actual initial FT/NFT differences that analysis reveals, we present ANOVA results only in juxtaposition to the results of the corresponding analysis of covariance (ANCOVA), as special "unadjusted" cases of the latter. A more complex procedure, ANCOVA adjusts observed contrasts to take at least partial account of initial FT/NFT inequalities, and is the simplest analytical procedure which reflects reasonably well the intended structure of the Follow Through experiment. Therefore, ANCOVA occupies a central place in our scheme: all others are either elaborations on ANCOVA or special limited cases of it.

Identifying ANCOVA as our fundamental mode of analysis does not solve all our analytical problems. We must still deal with such issues as the unit of analysis, the choice of outcome variables and covariables, the sensitivity of results to violations of assumptions, and the selection of an appropriate computational vehicle. Considerations arising from these problems motivate much of the organization of the remainder of this chapter.

2.1 Effects of Aggregation

The data permit us to ask many of our questions with respect to at least three distinct units ("levels") of analysis: child, class, and school. At the child level, we are in a position to investigate the full richness of interactions among characteristics of Sponsors, communities, classroom groups, teachers, and children; and we have enough data to permit detailed studies within Sponsors, regions, and types of children. At this level, on the other hand, we are most beset with the consequences of

measurement error. With no aggregation to average out error, our results reflect, for example, the underadjustment that results from the use of fallible covariates (Lord, 1967). Although adjustment procedures exist to correct for these biases in the case of a single imperfectly reliable covariate, appropriate adjustments have not yet been devised for the multi-covariate case. For this reason, the child level analyses have biases which we know exist but which we cannot eliminate because the covariate measurements are unreliable.

At the school level, the other extreme of the aggregation spectrum, a complementary set of advantages and drawbacks obtains. From thousands of children, we are reduced to a few hundred schools: still enough degrees of freedom for broad-brush studies, but insufficient to penetrate the fine structure of Follow Through. Measurement error, on the other hand, need not concern us at the school level: the stability of school means is much better than that of the individual child measurements that comprise them (Hannan, 1970). Class level analyses occupy a position between child and school analyses: aggregation of the data has decreased fallibility while also decreasing the level of detail of the analyses.

Since different substantive concerns motivate analyses at the three levels, we present the results of school, class, and child analyses, comparing them where appropriate. Where results are consistent for parallel questions across the three levels of aggregation, we have enhanced confidence that they represent the true effects. Where they are not consistent, we have some clues as to the sources of the discrepancies.

However valuable the cross-validation strategy, we have not permitted it to dominate our design of the analyses. Had our main purpose been to investigate the nature of aggregation biases, we would have limited the populations of the child level and class level studies to include only those children and classes which figured in the aggregations to school level. While such a study would have considerable methodological interest, we have foregone it in favor of the enhanced substantive interest that we feel is achieved by a more diverse set of analyses. We have, therefore, assembled the most inclusive analysis population available for the analyses at each level, without regard to the eligibility criteria of the other

levels. The discrepancies among results from level to level are confounded by (1) the effects of sampling; (2) the effects of aggregation; and (3) the intrinsic disadvantages of the various levels. Where patterns survive all these hurdles intact, our confidence in their reality grows. Where they are substantially dissimilar, we must look deeper into the data for the causes other than sampling effects.

3.0 PRESENTATION OF RESULTS

The results of the various analyses lend themselves to somewhat different modes of presentation, and the distinctive aspects of each display mode will be explained in the accompanying text. Enough elements occur in most of the effects profiles, on the other hand, to justify some introductory comments on their format and meaning. Figure VII-1 illustrates these common elements well. It displays the results of eight parallel analyses which yield measures of the "main effects" of Follow Through upon eight criterion measures, averaged across ten Sponsors. The statistics tabulated at the bottom of the figure document the derivation of the graphical display. They include:

- Adjusted and unadjusted values of the main effects, expressed in the units of the criterion variables. The unadjusted Absence effect of -1.070 (in the rightmost column) implies, for example, that FT school absence rates averaged a little more than a day lower than those of NFT schools. With regional and demographic inequalities taken into account by covariance adjustment, FT's advantage increases to 1.854 days. The adjusted effects are computed by ANCOVA and the unadjusted by ANOVA. We refer to these effects by the algebraic symbol B because we have selected, as a computational vehicle for our analyses, a particularly flexible multiple regression formulation which yields the "effects" of ANOVA and ANCOVA as raw score regression weights ("B weights"). These correspond to appropriately coded nominal predictor variables that reflect the desired analysis.
- The standard errors of the adjusted and unadjusted main effects, which we compute.
- t ratios as the ratios of the effects to their standard errors. As Monograph IV on methodological issues makes clear, one can regard the squares of these t ratios as scaled signal-to-noise ratios reflecting the extent to which the observed patterns of effects emerge clearly from the undifferentiated criterion variance that would otherwise seem unrelated to the treatments and

covariables. The t ratio magnitudes below 1.0, as a rule of thumb, characterize effects that one should not take too seriously. An absolute t ratio greater than 2.0 corresponds in the probabilistic analog, of course, to the $p < .05$ confidence level for two-tailed hypotheses. Our inferences are non-probabilistic of necessity; we report the t statistics without comment, merely to help the reader compare the relative importance of the effects that make up our profiles.

- The standard deviation of the criterion, reflecting the variability of each outcome variable from school to school within the population under analysis.[1] We introduce this statistic only so as to be able to compare the magnitudes of effects from one outcome variable to another.
- The effects expressed in criterion standard deviations, permitting the reader to say, for example, that FT's overall adjusted "effect" on the school average Wide Range Achievement Test (WRAT) scores amounts to 0.411 standard deviations. The arrows of the graphical displays correspond to these standardized effect measures, both adjusted (solid arrows) and unadjusted (dashed arrows).
- The number of schools used in the computation. This number, analogous to the sample size in a probabilistic analysis, has a strong influence, of course, on the standard errors of the effects and therefore on the t ratios. With large N's, it is possible to have large t's that correspond to totally uninteresting effects; with small N's, on the other hand, really important and revealing patterns can wash out in the noise. Our populations of two or three hundred schools strikes a balance: t ratios between 1.0 and 2.0 correspond to effects on the order of a quarter of a standard deviation, a conceptually interesting level of effect.

4.0 NOMINAL CODING SCHEMES

ANCOVA was chosen as a technique for evaluating the FT effects so as to adjust for the initial (previous to applying FT/NFT) differences among

---

[1] Rather than the simple total population standard deviation, we could have chosen, less conservatively, to use here the within-cell standard deviation pooled across all Sponsor x treatment combinations, thus eliminating the between-cell component of variability and altering the numerical size of the standardized effect. The principal consequence of such a choice would have been to spread out somewhat the scale of the graphical display of Figure VII-1. The change would also reveal, to be sure, variations in the pattern of intercell variability from outcome to outcome, but we doubt that these modifications would alter significantly the overall patterns that the displays are designed to reveal.

the participants (children, classes, schools, teachers, parents, etc.). However, since the number of participants in various groups (Sponsors, regions, etc.) were unequal--either because of the initial design, or because the students (and their parents) moved, or because of missing or inadequate data--and also since many of our covariates were not continuous, the choice of available computer packages for performing the analysis was either limited or nonexistent. On the other hand, a variety of standard computer packages (e.g., Statistical Package for the Social Sciences [SPSS]), are available for conducting analysis of regression studies. Thus, instead of developing our own custom-made ANCOVA packages (a costly alternative), we decided to use the SPSS regression package to conduct our ANCOVA studies. The use of nominal coding schemes in multiple regression for performing analysis of covariance (in fact, any general linear hypotheses modelling) is not a novelty to mathematical statisticians. For example, Scheffé (1959) has used the general linear model to develop the foundations of ANOVA. It seems, however, to have been largely ignored by researchers in social sciences. For example, Jacob Cohen's 1968 paper marks one of the first instances of gainfully employing this technique in psychological research (Cohen, 1968). Our evaluation of Follow Through by these methods for performing analysis of covariance is an example of this new trend in social research. For this reason, we have developed a somewhat detailed discussion of the methodological issues associated with using this technique in Monograph IV. A brief summary of our method follows.

4.1 Classical Approach

In the classical one-way ANOVA model, there is one research factor whose effect is studied at many (say k) levels. The analytical model assumes that each observed value of an outcome measure (say Y) can be partitioned into three terms: a term representing the general mean effect (say M); another representing the eccentricity (the "effect") associated with the particular level of the research factor (say A); and the third representing a "normally distributed" error (say e). If there are n_i

observations at the i^{th} level of the research factor ($i = 1, 2, \ldots, k$), then the value of the j^{th} observation ($j = 1, 2, \ldots, n_i$) is represented by:

$$(1) \qquad Y_{i,j} = M + A_i + e_{i,j}.$$

The classical analytical schemes proceed to estimate the parameters M, A_i, and σ^2--the latter being the variance of the error terms. One computes the "within sum of squares" (WSS) and "between sum of squares" (BSS), and proceeds to test the null hypothesis that the factor levels do not have statistically significant effects (i.e., that $A_i = 0$). Since the research factor and the factor levels are generally chosen to demonstrate the differences among levels, the data analyst usually arrives at a not-very-surprising finding that the null hypothesis is untenable. Surprisingly few social researchers then proceed to follow up on this to (1) detect which levels are significantly apart from each other and (2) test the strength of the relationship between the outcome measure and the factor level (Hays, 1963).

4.2 <u>Rationale for Our Scheme</u>

Instead of the classical methods of data analysis, we have chosen to convert ANOVA into analysis of regression by employing the nominal coding scheme. The advantages are many: (1) <u>the method is exact</u>, so that the classical method and ours would come to the same set of conclusions regarding significance of the factor levels; (2) <u>the strength of relationship</u>, as a suitably chosen correlation coefficient, is always computed; (3) <u>the method is easily generalizable</u> to more than one factor and their interactions; (4) <u>it is well suited</u> to describe an ANCOVA model wherein some factors are categorical and others vary continuously; and (5) <u>it allows us to consider different interpretations of the main effects</u>.

4.3 <u>An Example</u>

Consider, for example, the investigation of variations in the WRAT scores as a function of sponsorship. If the observations are divided into

k groups (defined by the Sponsor associated with the group), any set of k - 1 linearly independent predictors uniquely represents the sponsorship as a factor; the one which automatically yields the classical main effects (A_i of the earlier discussion) is described here. Let the first predictor equal 1 for those associated with the first Sponsor, -1 for those with Sponsor k, and 0 otherwise. Each of the k - 1 predictors is chosen similarly. The WRAT scores of each participant and the concomitant values of the nominal predictors are introduced in a regression equation. Corresponding B weights (the coefficients in the best fitting equation) equal A_1 through A_{k-1}; and A_k equals the negative of the sum of all B weights. The mean effect (M) is given by the constant term. The variance of the error terms is the residual variance exhibited by the standard regression packages. Finally, the classical ANOVA F ratio equals the F ratio for testing the significance of the regression.

4.4 Generalizations

This is one of the nominal coding schemes. Cohen (1968) identifies it by the name "Effects Coding Scheme." This and other nominal coding schemes are described more fully in Monograph IV. When there are several research factors, each is coded by an appropriate nominal coding scheme; the observation groups are identified uniquely by the predictor values; the interactions amongst the research factors (where appropriate) are always represented by the product of corresponding main effects predictors; and the multiple R^2 is always used for testing the null hypotheses of no overall effect. This method also allows one to estimate the explanatory power of each factor (and in fact of each predictor) in the model--something not usually done by classical ANOVA users.

The extension to ANCOVA is relatively straightforward. The predictors corresponding to each covariate (or each aspect of a covariate) are introduced along with those representing the research factors. The statistical equality provided by a covariate (set) is checked by testing the strength of covariate x research factor interactions--a significant interaction indicating the inappropriateness of the corresponding covariate.

If the values of covariates for some participants are missing, the randomness of the missing values can be examined by specifically introducing a "missingness" predictor and testing it for significance.

5.0 THE CHOICE OF EFFECTS

In the classical applications of ANOVA or ANCOVA, the "effect" of a level of a research factor is envisioned as the difference between the mean of outcome measures at that level of the research factor, and the "grand mean" of such measures across all levels. This is most appropriate if the levels represent variations or graduations of some "treatment." It is not appropriate if one of the levels is interpretable as a "control," i.e., lack of a treatment. The difference here is not one of amount but rather of kind. Under these circumstances, it is more appropriate to envision the effect to be the difference between the mean of the outcome measures at a level of the research factor, and the corresponding mean for the level identified as the control. Our method provides a "natural" coding scheme to derive such effects.

Consider, for example, the coding scheme for Sponsors discussed earlier. If one of the Sponsors (say Sponsor k) was in fact a "control," we can capitalize on this knowledge as follows. Define the first nominal predictor to take on the value 1 for observations with Sponsor 1, and 0 otherwise. Similarly, let the second predictor equal 1 for those with Sponsor 2, and 0 otherwise. One needs, as before, only k - 1 such predictors. The control group needs no special predictor of its own: it is identified by being zero-coded on <u>all</u> predictors. The F ratio continues to have the same meaning: its value measures the adequacy of the general model. However, the B weights now measure the desired difference between the factor level means and the control mean. The constant of the regression equation now equals the mean for the control group.

The flexibility afforded by using different schemes is desirable; it allows us to concentrate on problem formulation rather than on mundane computations to bring into focus the appropriate effects. The reader, on the other hand, must be cautious while reading some of the substantive

chapters of this report: he cannot assume that the reported effects have classical meaning. He must exercise caution whenever the study examines the effects of more than one factor and corresponding interactions: some factors have been coded by the "effects coding" scheme, others employ the "control coding" scheme. Please refer to Monograph IV for a more detailed discussion of this "linguistic problem," particularly while interpreting the interactions.

After a brief description of several studies which parallel those pertaining to children we will proceed with a discussion of a number of studies at the school level of analysis employing the analytic tools described above. Although the designs of the studies are more complex than the expository cases presented, the principles are identical.

CHAPTER VI

OVERVIEW OF PARALLEL STUDIES

The content of this volume is concerned with the analyses of pupil outcome data as a function of membership in a particular Sponsor's program. However, information on other possible contributors to these analyses, which is not included in the covariable set, is also available in Volume I-B. Monograph I of that volume discusses an investigation of the differences among parents or factors which may have an impact upon children's achievement scores. Monograph II addresses a similar issue with respect to teachers: Are there differences among teachers or a variety of teacher characteristics? Monograph III presents further evidence that there may be critical differences among school districts or dimensions of implementation of Sponsor programs. Whereas none of these three studies have been merged with pupil data, the results indicate that differences occur among Sponsors which may help to clarify identified differences between their FT and NFT groups.

Summaries of these studies are included in this chapter to indicate to the reader some properties of these parallel investigations.

1.0 SUMMARY: PARENT STUDIES

One of the basic tenets of the Follow Through Program is that children's educational progress is influenced by several aspects of their environment. Thus, the parents' attitudes and behaviors and possibly also their socioeconomic status are seen as potential mediators in their children's educational success. For this reason, Follow Through Program Guidelines encourage and mandate parent involvement, and a number of our research concerns center on identifying the significant relationships among parent variables in order to eventually determine to what extent these variables and which of them may be related to children's success.

The initial parent studies basically include three sets of analyses. The first two are parallel but separate examinations of Cohort III (kindergarten) and Cohort I (third grade) parents; the third is a series of analyses designed to identify interrelationships among several sets of analyses.

In both Cohorts, the overall FT/NFT comparison of family income and mother's education level suggest that FT families have lower incomes and are less well educated than NFT families. While this difference may be an artifact of sample selection, the case can also be made that FT appears to be reaching the lower socioeconomic groups for whom it was intended. In both Cohorts I and III, although Sponsor FT/NFT contrasts almost always maintain the same direction as overall, there is considerable variability Sponsor by Sponsor. These local differences may reflect important contextual differences in the relationships Sponsors have with their clients, and it could be that simply adjusting the pupil outcomes by a socioeconomic covariate set which include only income and educational level may not begin to do justice to the real differences among Sponsors. The third demographic variable studies, family mobility in both Cohorts, showed surprisingly little variability between FT and NFT either overall, or within Sponsor. It may be that there are differences in mobility in the individual from site to site which disappear when the data are aggregated by Sponsor. In any case, however, if patterns of mobility change over time, this variable may provide an indicator of the nature of communities useful in future studies.

In general, we found in Cohort III that FT parents were more involved with their child's schooling on three measures of involvement, and more satisfied with academic success than were NFT parents. Since parents were interviewed as late as November in some sites, this trend favoring FT may reflect an early program effect, or may be the result of initial differences in the two groups (FT and NFT) of parents.

FT parents in Cohort I reportedly interact more with their children's schools and are more satisfied with their children's affective growth than are NFT parents. In spite of these differences in both FT and NFT, parent satisfaction is relatively high, and in both cases involvement is moderate. In this Cohort Sponsor variability occurs in only one case: satisfaction with affective growth.

The examination of Cohort I data for potential mediators showed some complex relationships, and differing Sponsor effects that are worthy of further investigation. The more parents interact with their child, the more likely they are to be satisfied with his affective growth; the higher their income, the more likely they are to interact with schools. In addition, there is an overall positive relationship between perception of

the school's receptivity and satisfaction with affective growth. Perception of receptivity shows no relationship with parent-school interaction.

These initial explorations have demonstrated the complexity of the FT data. The Sponsor variability observed--although not terribly frequent--strongly suggests that Sponsors have differing effects on different types of parents. Overall, however, it must be remembered that FT/NFT differences consistently favor FT.

2.0 SUMMARIES: TEACHER STUDIES

The teacher is the person who must translate a Sponsor's theoretical approach into classroom experiences for children. While we have not yet merged teacher and pupil data, the teacher studies shed some light on variations in a number of important teacher characteristics including: (1) personal and professional background; (2) training in basic Sponsor philosophy; (3) values, attitudes, and reported behaviors; and (4) satisfaction and perceived faithfulness to the Sponsor's approach. The teacher studies also explore the relationship of teacher background to the other teacher characteristics, in order to determine whether or not Sponsor delivery and/or implementation are mediated by the characteristics teachers bring with them to the program. A group of 1122 FT and NFT teachers from kindergarten through third grade served as the source of teacher data. All information was drawn from a teacher questionnaire administered in the spring of 1972.

FT teachers, on the average, were found to be slightly younger and less experienced than NFT teachers, although both groups had been teaching for several years. In addition, the FT group had slightly lower salaries than the NFT group. The two differed little in their educational attainment, with the vast majority of teachers having earned advanced credits or degrees. The FT group had slightly more minority teachers, as well. In addition to these overall differences, there were FT/NFT differences within Sponsor and grade level. Many of these differences appeared to be related to community size and region, with teachers in non-Southern, large cities being more apt to be highly educated, better paid, and non-white.

Sponsor variations were also found in both the amount and focus of the training FT teachers reported receiving. Training was classified in three

areas: structure, child-centeredness, and working with parents and aides. Some Sponsors, like Sponsor 12, appeared to provide a relatively large amount of training in all three areas. Others, like Sponsors 2, 5, 9, and 14 appeared to provide relatively little training, although what training was given reflected Sponsor philosophy. Still others, like Sponsors 3, 7, 8, 10, and 11 provided highly differentiated training programs, extending a great deal of training in those areas related to basic principles.

It was not possible to group Sponsors into categorical types on the basis of the delivery of training. Even those Sponsors which are most often linked together--Sponsors 7 and 8--had very different training profiles. In addition, training was found to vary by the size of the community in which the program was located. The bigger the city, the less training teachers reported receiving in child-centeredness. There was also a tendency for teachers in rural areas to report receiving less training from some Sponsors.

Turning to teacher values, attitudes, and behaviors, both FT/NFT and grade level differences were found. FT kindergarten teachers differed from NFT kindergarten teachers in their attitudes and reported behaviors toward parents, with FT kindergarten teachers much more apt to value meeting with parents and to visit pupil homes. Kindergarten teachers, in general, were more apt to have positive attitudes toward parent involvement than teachers at higher grades. They were also more apt to value a child-centered approach to education than a structured approach.

FT/NFT contrasts within Sponsors were also explored. Differences were examined in teacher values toward: parent-community orientation; social skills development; and structured/academic vs. child-centered orientation, as well as frequency of teacher visits to pupil homes. Once again, Sponsor differences were found.

While some of these differences reflected Sponsor's theoretical orientations, others did not. Nor were the findings completely consistent across grade levels or communities. It was pointed out that in many instances failure to find a significant FT/NFT contrast did not represent a failing on the part of the Sponsor but rather an impartial NFT group. Sponsor approaches are not chosen in a vacuum; the initial choice of a Sponsor's approach may reflect a basic community orientation toward that

approach. In addition, following three years of implementation, a Sponsor's approach may well be diffused throughout a school system. For either or both of these reasons, it appears that some Sponsors have NFT groups with values closely reflecting their own philosophical approach. These similarities must be considered in examining implementation questions and in understanding FT/NFT pupil contrasts, as well.

Finally, despite variations among Sponsors and communities, it was found that FT teachers overall were extremely satisfied with the program and perceived themselves as faithful to their Sponsor's approach. Satisfaction and perceived fidelity were strongest in the kindergarten, Cohort III groups.

3.0 SUMMARY: IMPLEMENTATION STUDY

This study was designed to examine several questions related to the implementation of Sponsor programs in the schools. These included: (1) the manner in which Sponsors were selected or assigned to schools; (2) the relationships between Sponsors and Local Education Agencies (LEAs); and (3) LEA problems, idiosyncratic characteristics of staff members, and non-program-related events which might affect program delivery or implementation.

Two Big City sites were chosen in which to explore these questions: Philadelphia and New York. These sites were selected so that Sponsors could be examined in a relatively homogeneous context. Data were completed primarily by means of semi-structured interviews with the FT director or Program Coordinator at these sites.

The implementation study highlights the non-random nature in which Sponsors were assigned/selected, both across and within sites. In Philadelphia, schools were primarily assigned to FT Sponsors by the District Superintendents. Moreover, concern for city-wide experimental design limited the choices available to Superintendents. In New York, Sponsors were selected by schools and parent representatives in a relatively free setting. Here, too, choices were limited, however; once a Sponsor was chosen by one school, it could not be selected by another.

In both sites, variations in the selection/assignment procedures as well as in Sponsor communication led to differences in the

responsiveness of schools to the Sponsor's approach. More specifically, differences in responsiveness were found among district and school administrators, teaching staff members, and parents. In some instances, the various groups were all in favor of the Sponsor's approach, in others none were, and in still others, there was conflict among these groups. Many of these differences were found to be associated with differing value systems among the several groups. However, staff turnover and parent mobility also differentially affected lines of communication between Sponsors and schools.

Finally, LEA problems and non-program-related events were found to affect program implementation differentially. Teacher strikes, decentralization plans, fund cuts, and the redistribution of teachers due to reductions in school enrollment were some of the LEA problems affecting implementation. Non-program-related events included changing neighborhoods, and school construction. In most cases, these events disrupted the school program; in others they served as facilitators, strengthening school-community ties.

In conclusion, the implementation data collected to date highlight the fact that a Sponsor's program cannot be examined independently of the manner and context in which it is implemented. The wide variations in the ways in which Sponsors are selected, in program delivery, in the manner in which the programs are received by staff and community members may have a differential impact on the extent to which a program is implemented. These initial explorations will guide future data collection and analysis; for the present they serve as an important backdrop against which the pupil outcomes should be viewed.

CHAPTER VII

SUMMARY: FT/NFT CONTRASTS

We now turn to the first two goals as set forth in the introduction:

1. To identify the FT/NFT contrasts in posttest scores on achievement, motivation, and absence measures for each Sponsor and all Sponsors combined.

2. To identify the FT/NFT contrasts in posttest scores for each Sponsor associated with a sample of children in Big Cities.

We first approach Goal 1 by looking at the FT/NFT contrasts in posttest scores globally across all Sponsors and individually for each Sponsor at the school level of analysis. These contrasts are based on the one year data for the Cohort III kindergarten children; the analytic subset does not include the Big City schools.

In the next study these Big City schools (located in New York, Philadelphia, and Chicago) are included in the sample to enable us to indirectly identify the Big City effects as specified in Goal 2.

Additional information concerning the respective roles of socio-economic status and time of testing is presented next to amplify these FT/NFT contrasts.

The fourth section of this chapter presents FT/NFT contrasts on posttest scores for each Sponsor not only at the school level but also at the class and child levels of analysis. These contrasts are juxtaposed in an attempt to further describe Sponsor effects in a variety of contexts. These Sponsor vignettes are based on the total sample (including the Big City sites).

This chapter closes with a brief look at the Cohort I findings. First we examine a three-year longitudinal study of Cohort I entering first graders; then we compare Cohort I and Cohort III kindergartners in an attempt to see if school level effects have changed from Cohort I to Cohort III. Although these studies do address the issue of FT/NFT contrasts for each Sponsor, they are important primarily as prototypes of future analyses when data are more complete.

1.0 THE ONE YEAR KINDERGARTEN STUDY, SCHOOL LEVEL

1.1.0 INTRODUCTION

As an initial step in the evaluation of Follow Through we could ask in the One-Year Kindergarten Study the question: "Is there a global FT effect?" That is to say, when we merge all of the various Sponsors into a single group and treat FT as an undifferentiated program, is there a detectable "Follow Through" effect?

The meaning of the answer to this question, however, has to be elaborated, because "Follow Through" in reality consists of a variety of different programs and approaches. It is perfectly possible, for instance, that a positive overall "FT effect" may be due to large effects on the part of a very few Sponsors, while the rest are having no effect. Or we could find "no overall FT effect" when in fact some Sponsors are having a strong positive effect while other Sponsors are having a strong negative effect, thus canceling each other out in an overall assessment of the FT effect. The major thrust of our analyses therefore lies in asking, and answering, the questions which provide an elaboration of the meaning of an overall "Follow Through effect": What are the particular effects of particular Sponsors under particular conditions?

1.2.0 METHOD

1.2.1 Design

The major function of the design is to identify the nature and extent of the contribution of the several models (Sponsors) to the overall FT effect. In order to accomplish this, it is necessary to adjust Sponsors for initial differences on a variety of scores. Next, it is necessary to consider the pattern of Sponsor contributions across a variety of outcomes. Finally it is necessary to attempt an adjustment of the original mismatch between FT/NFT groups. The procedure of choice is an analysis of covariance: Sponsors, and their FT/NFT groups are examined as independent variables; sets of covariables are utilized as adjusting variables for Sponsor mismatches; a child pretest measure is used to adjust for initial differences among children; and several achievement and motivational variables are used separately as criterion measures.

1.2.2 Analytic Subset

The study is based on the kindergarten class of 1971-1972. A detailed explanation of why this cohort of children was selected has been provided in Chapter II. Also included in that earlier chapter are the reasons for focusing on the school level of analysis.

This study is based on a subset of 251 schools (137 FT; 114 NFT) distributed across ten Sponsors as displayed in Table VII - 1. The selection criteria required each school to have data (i.e., school means and standard deviations) on all the variables defined in the following section. None of the Big City schools are included in this analytic subset. They will be considered in Section 2.0 of this Chapter.

1.2.3 The Variables

Twenty-one variables are included in these school level analyses: eight criterion or outcome measures, two indicators of "treatment", and eleven covariables, aggregated where necessary to school level.

Criterion Measures

Four measures of achievement, one of achievement motivation, two of Locus of Control, and one of school attendance comprise the present battery of criterion measures: aspects of a child's life that the Sponsors aim to influence, to varying degrees.

- Wide Range Achievement Test (WRAT), in a version shortened and adapted especially for Follow Through, and administered in Spring, 1972.

- Metropolitan Achievement Test (MAT), the raw scores of three separate subtests:

  - - Listening to Sounds
  - - Reading
  - - Arithmetic

- Gumpgookies, a measure of achievement motivation

- Locus of Control, two subscores:

TABLE VII - 1

Distribution of the FT and NFT Schools
Across Ten Sponsors For the Study of
One Year Kindergarten Effects

SPONSORS (by code number)

	2	3	5	7	8	9	10	11	12	14	TOTALS
FT	29	21	12	11	16	8	15	10	9	6	137
NFT	20	20	10	9	12	9	9	9	11	5	114
TOTALS	49	41	22	20	28	17	24	19	20	11	251

-- Internal Positive, reflecting the extent to which a child believes that he is responsible for good events in his life.
-- Internal Negative, reflecting the extent to which he believes that he is responsible for bad events.

- Absence, the number of days missed during the school year.

Treatment Variables

Two nominal treatment variables, Sponsor (with ten "levels") and FT/NFT (with two levels) define twenty "treatment groups". For the sake of computational convenience and flexibility, we cast the analyses of covariance into a mathematically-equivalent multiple-regression format as described in Chapter V. The initial research questions dictate two treatment-variable configurations:

- The Nested Design, equivalent to a 2 x 10 analysis of covariance with FT nested within the Sponsor groups. Table VII-2 displays the corresponding treatment variables coding, designed so that the raw-score regression coefficients for the ten predictors will be numerically equal to the effects of FT within each of the ten Sponsors, measured in the units of the criterion variable. The standard errors of those regression coefficients become measures of the "significance" of the effects of interest. Those tables depicting a Sponsor's FT effect are displaying results of an ANCOVA using a nested design.
- The Factorial Design, equivalent to a 2 x 10 factorial analysis of covariance. Table VII-3 displays the required coding scheme. The raw-score regression coefficient of the predictor labeled "Main" is the main effect of FT, the simple mean of the ten Sponsor effects computed under the corresponding Nested Design. This analysis is done for the sake of the auxiliary statistics (standard errors, proportions of variance) associated with the main effects that do not follow directly from the nested analog as does the effect's magnitude itself. Those tables depicting the overall main effect of FT are displaying results of an ANCOVA using a factorial design.

Covariables

The interpretation of our results depends heavily upon the nature and reliability of the set of variables used to take account of initial FT/NFT group mismatches. Chapter IV discusses the way in which the eleven covariables were selected and checked to be sure they would function properly in their role. The final list is as follows:

TABLE VII - 2

Predictor Coding Scheme for the Nested Analysis of One-Year Kindergarten Effects

Sponsor	FT/NFT Status	10 FT-NFT Effects Within Sponsors										9 Contrasts Among Sets of Sponsors								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	FT	.5	0	0	0	0	0	0	0	0	0	1	-4	4	0	0	0	0	0	0
	NFT	-.5	0	0	0	0	0	0	0	0	0	1	-4	4	0	0	0	0	0	0
3	FT	0	.5	0	0	0	0	0	0	0	0	1	1	-1	1	1	1	0	0	0
	NFT	0	-.5	0	0	0	0	0	0	0	0	1	1	-1	1	1	1	0	0	0
5	FT	0	0	.5	0	0	0	0	0	0	0	1	1	-1	1	1	-1	0	0	0
	NFT	0	0	-.5	0	0	0	0	0	0	0	1	1	-1	1	1	-1	0	0	0
7	FT	0	0	0	.5	0	0	0	0	0	0	1	1	-1	-1	-1	0	1	0	0
	NFT	0	0	0	-.5	0	0	0	0	0	0	1	1	-1	-1	-1	0	1	0	0
8	FT	0	0	0	0	.5	0	0	0	0	0	1	1	-1	-1	-1	0	-1	0	0
	NFT	0	0	0	0	-.5	0	0	0	0	0	1	1	-1	-1	-1	0	-1	0	0
9	FT	0	0	0	0	0	.5	0	0	0	0	-1	-4	-4	0	0	0	0	0	0
	NFT	0	0	0	0	0	-.5	0	0	0	0	-1	-4	-4	0	0	0	0	0	0
10	FT	0	0	0	0	0	0	.5	0	0	0	-1	1	1	1	-1	0	0	1	0
	NFT	0	0	0	0	0	0	-.5	0	0	0	-1	1	1	1	-1	0	0	1	0
11	FT	0	0	0	0	0	0	0	.5	0	0	-1	1	1	1	-1	0	0	-1	0
	NFT	0	0	0	0	0	0	0	-.5	0	0	-1	1	1	1	-1	0	0	-1	0
12	FT	0	0	0	0	0	0	0	0	.5	0	-1	1	1	-1	1	0	0	0	1
	NFT	0	0	0	0	0	0	0	0	-.5	0	-1	1	1	-1	1	0	0	0	1
14	FT	0	0	0	0	0	0	0	0	0	.5	-1	1	1	-1	1	0	0	0	-1
	NFT	0	0	0	0	0	0	0	0	0	-.5	-1	1	1	-1	1	0	0	0	-1

TABLE VII - 3

Contrast Coding Scheme for the Factorial Analysis of One-Year Kindergarten Effects

Sponsor	FT-NFT Status	9 Contrasts Among Sponsors									Main FT-NFT Effect	9 Sponsor-Effect Contrasts								
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	FT	1	0	0	0	0	0	0	0	0	.5	.5	0	0	0	0	0	0	0	0
	NFT	1	0	0	0	0	0	0	0	0	-.5	-.5	0	0	0	0	0	0	0	0
3	FT	0	1	0	0	0	0	0	0	0	.5	0	.5	0	0	0	0	0	0	0
	NFT	0	1	0	0	0	0	0	0	0	-.5	0	-.5	0	0	0	0	0	0	0
5	FT	0	0	1	0	0	0	0	0	0	.5	0	0	.5	0	0	0	0	0	0
	NFT	0	0	1	0	0	0	0	0	0	-.5	0	0	-.5	0	0	0	0	0	0
7	FT	0	0	0	1	0	0	0	0	0	.5	0	0	0	.5	0	0	0	0	0
	NFT	0	0	0	1	0	0	0	0	0	-.5	0	0	0	-.5	0	0	0	0	0
8	FT	0	0	0	0	1	0	0	0	0	.5	0	0	0	0	.5	0	0	0	0
	NFT	0	0	0	0	1	0	0	0	0	-.5	0	0	0	0	-.5	0	0	0	0
9	FT	0	0	0	0	0	1	0	0	0	.5	0	0	0	0	0	.5	0	0	0
	NFT	0	0	0	0	0	1	0	0	0	-.5	0	0	0	0	0	-.5	0	0	0
10	FT	0	0	0	0	0	0	1	0	0	.5	0	0	0	0	0	0	.5	0	0
	NFT	0	0	0	0	0	0	1	0	0	-.5	0	0	0	0	0	0	-.5	0	0
11	FT	0	0	0	0	0	0	0	1	0	.5	0	0	0	0	0	0	0	.5	0
	NFT	0	0	0	0	0	0	0	1	0	-.5	0	0	0	0	0	0	0	-.5	0
12	FT	0	0	0	0	0	0	0	0	1	.5	0	0	0	0	0	0	0	0	.5
	NFT	0	0	0	0	0	0	0	0	1	-.5	0	0	0	0	0	0	0	0	-.5
14	FT	-1	-1	-1	-1	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5	-.5
	NFT	-1	-1	-1	-1	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5	.5	.5	.5	.5

- Wide Range Achievement Tests (WRAT), administered as an achievement pretest in Fall, 1971, and expressed as a school mean of child scores in participating FT or NFT classes.

- Percentage of Black pupils in the school's study populations.

- Percentage of minority pupils, including Indians, Orientals, and Spanish-speaking children as well as Blacks.

- Years at current address, according to information provided in parent interviews, averaged to school level.

- Adjusted income level, the school average of a composite prosperity index which incorporates parent-reported income, family size, and whether or not the family is located in a rural area.

- Mothers' Education: for each school, the percentage of mothers reporting at least a high school education.

- Parent-school receptivity: a composite index, discussed in detail in Monograph I, reflecting the extent to which parents perceive that their child's school welcomes parental participation.

- Western region: 1 if the school is located in the West; 0 otherwise.

- Southern region: 1 if the school is located in the South; 0 otherwise.[1]

- Metropolitan: 1 if the school is located within a Standard Metropolitan Statistical Area (SMSA); 0 otherwise.[1]

- Middle-sized cities: 1 if the school is located in a metropolitan community whose population falls between 50,000 and 200,000; 0 otherwise.

These covariables measure a number of the ways in which schools differed at the beginning of an FT "treatment" so as to cloud the interpretation of post-treatment differences. Monograph III on implementation provides some insight into many other covariables which we would like to have used, particularly with relation to the degree, nature, and timing of Sponsor model implementation. The eleven covariables re-

---

[1] See Chapter IV for report of justification for inclusion of these two regional variables and for the exclusion of other regional variables from this covariable set.

present factors which would operate even if all models were delivered in all schools with comparable fidelity, and so they establish as much comparability among schools as the circumstances permit.

1.2.4 The Power of the Analyses[2]

Before displaying patterns of effects, it is appropriate to provide some statistical evidence of the ability of our analyses to discern meaningful patterns. If FT and sponsorship account for only negligible portions of the variability in our outcome measures, then the observed "effects" are trivial noise and there would be no value to displaying and discussing them. If, on the other hand, the FT "signal" pierces unmistakeably through the ambient "noise", then we can seek to understand the causal basis of the observed patterns.

Tables A VII-1 and A VII-2 of the Appendix display the complete partition of the variance of our eight criterion variables that the nested and factorial analyses accomplish. The accompanying tables of F-statistics (Tables A VII-3 and A VII-4) translate this purely descriptive partition into statements of the statistical significance of the observed contrasts between FT and NFT, both within and across Sponsors. Monograph IV on Methodology explains the logical and mathematical justification for our computation of these statistics and the interpretation we place on them.

1.3.0 PATTERNS OF EFFECTS

1.3.1 Main Effects

Figure VII-1 displays, both numerically and graphically, the results of eight parallel factorial analyses: the "main effects" of FT upon the eight criterion measures, averaged across the ten Sponsors. Refer to Chapter V on methodological issues for an explanation of the information found in these figures.

The main effects illustrated in Figure VII-1 permit us to make an affirmative response to the question, "Does FT do any good?" The overall covariance-adjusted effects of FT for the eight outcomes at the school level:

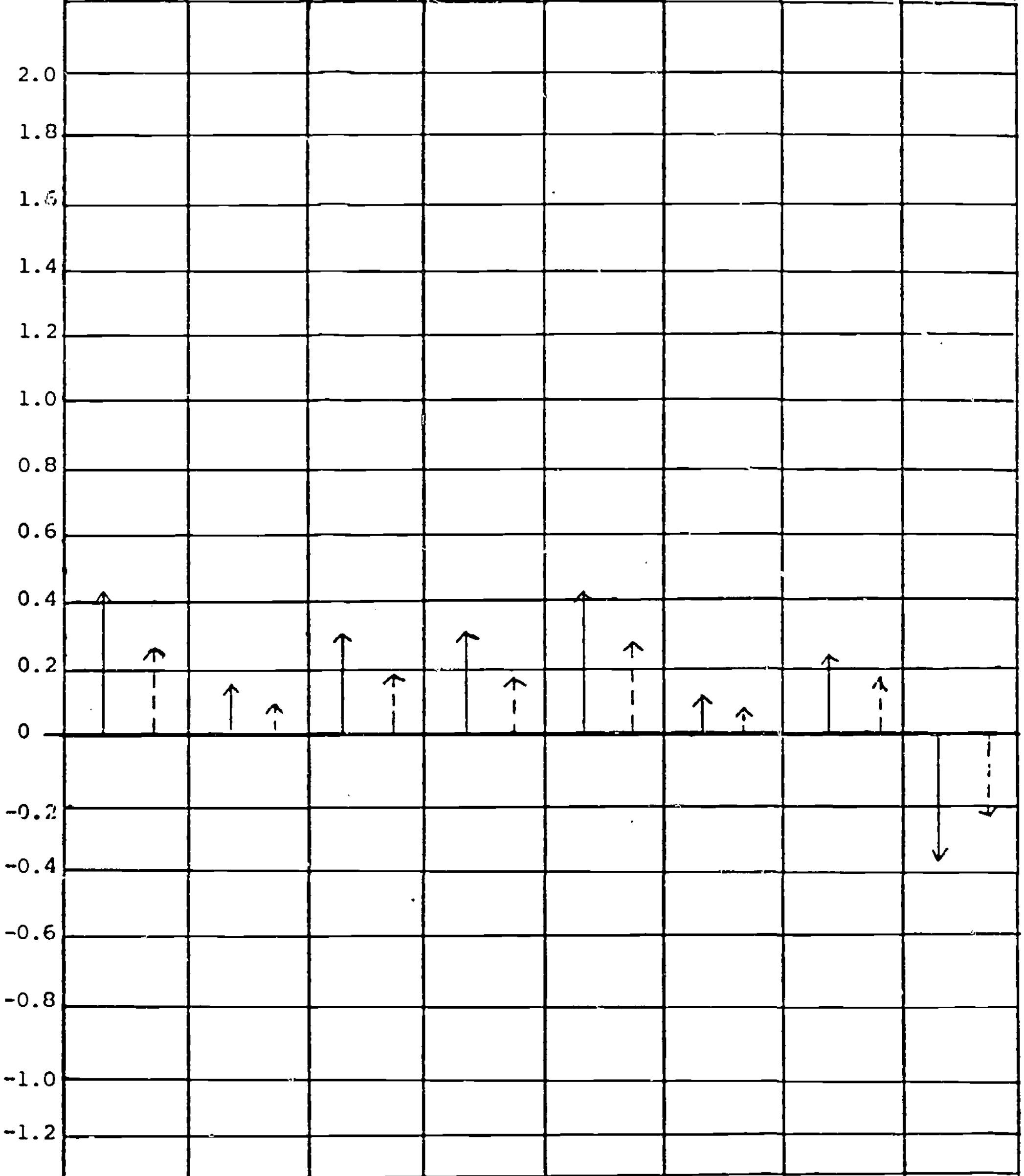
[2] The four tables referenced here display the information for this study as well as that which follows in Section 2.0.

Figure VII - 1

FOLLOW THROUGH EFFECTS PROFILE FOR MAIN EFFECT

10-Sponsor School Population, Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.358	0.643	0.731	0.961	4.656	0.064	0.079	-1.854
	Unadj.	0.220	0.341	0.471	0.571	3.043	0.022	0.061	-1.070
SE = Standard Error of B	Adjusted	0.087	0.410	0.285	0.359	1.349	0.070	0.053	0.628
	Unadj.	0.109	0.445	0.309	0.382	1.303	0.075	0.046	0.648
t = B/SE = "Significance" Statistic	Adjusted	4.115	1.569	2.567	2.677	3.452	0.917	1.493	-2.952
	Unadj.	2.018	0.768	1.526	1.496	2.335	0.295	1.329	-1.653
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.411	0.172	0.300	0.304	0.413	0.117	0.238	-0.370
	Unadj.	0.253	0.091	0.193	0.181	0.270	0.040	0.183	-0.213
N = Number of Schools in Computation	FT	137	137	137	137	137	137	137	137
	NFT	114	114	114	114	114	114	114	114

- are all in the desired direction (positive for all outcomes except Absence),
- emerge substantially from the noise, and
- are of the order of a quarter of standard deviation in magnitude (ranging from 0.117 to 0.413 standard deviations).

The magnitudes of the adjusted effects are all around a tenth of a standard deviation greater than those of the corresponding unajusted effects, reflecting the fact that FT children started out behind NFT children. The increase in adjusted effects over unadjusted effects makes it clear that simple unadjusted comparisons would do the initially lower-performing group an injustice and that covariance adjustment makes a more equitable comparison.

1.3.2 Sponsor Effects

With our overall main effects in hand, we are now ready to penetrate to the first level of fine structure: the Sponsors. Is there Sponsor variation within the overall FT effect?

Figures VII-2 through VII-11 leave no room for doubt as to the answer: Sponsor diversity is great. Sponsors 7 and 8 (University of Oregon and University of Kansas), with their strong positive achievement effects, have rather similar patterns, but their achievement motivation (Gumpgookies) and Locus of Control patterns differ markedly. University of Oregon (Sponsor 7), for example, is the only Sponsor with no relative effects on achievement motivation; he makes a strong showing, on the other hand, with respect to negative (but not positive) Locus of Control. Sponsor 2's (Far West's) effects seem to concentrate in reading and achievement motivation; Sponsor 10 (University of Florida) adds Locus of Control to this list; and Sponsor 12 (University of Pittsburgh) achieves mainly in arithmetic, achievement motivation, and Locus of Control. If there is an "average Sponsor" it is Sponsor 9 (High/Scope): his pattern looks very much like the overall main effects pattern, but stronger. Sponsors 3, 5, and 14 show mixed positive and negative effects, of which Sponsor 14's (SEDL's) sizable negative achievement motivation effect is the least typical. Except for a good-sized negative (i.e., favorable) effect on Absence, Sponsor 11

Figure VII - 2

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 2

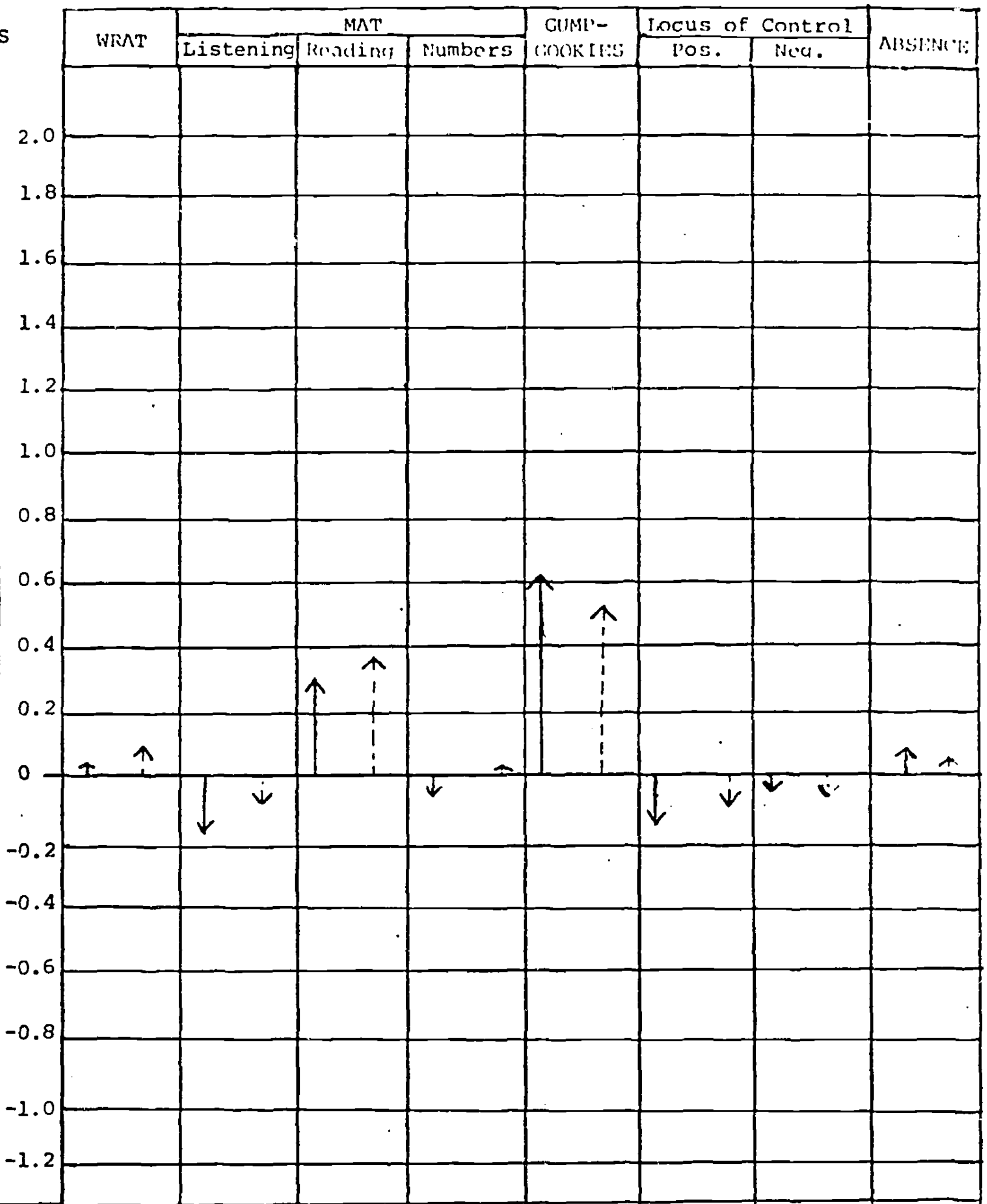
Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect

Unadj. Effect



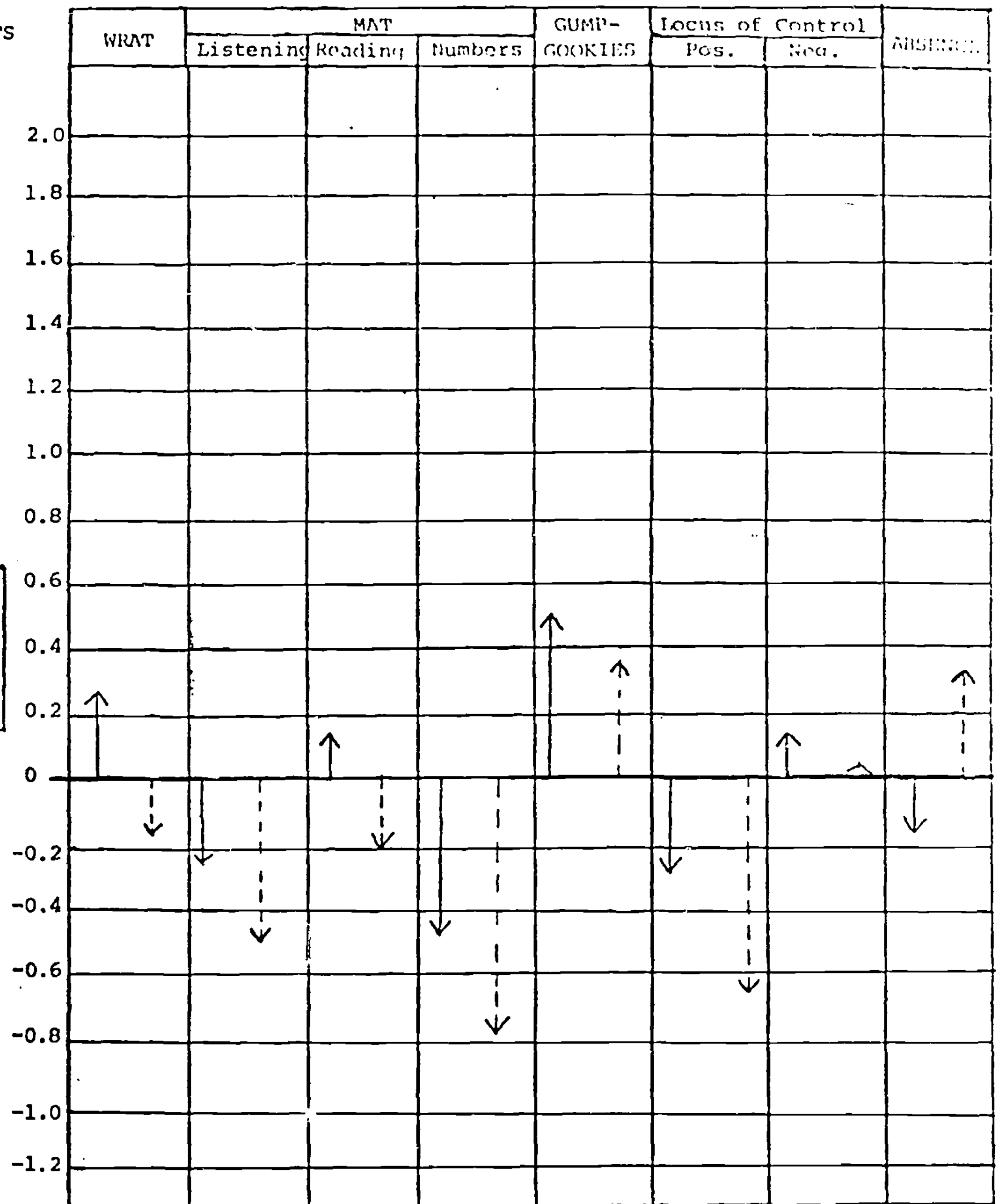
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.020	-0.683	0.690	-0.201	6.847	-0.088	-0.016	0.415
	Unadj.	0.085	-0.301	0.938	0.039	6.086	-0.041	-0.022	0.219
SE = Standard Error of B	Adjusted	0.158	0.743	0.516	0.650	2.444	0.126	0.096	1.138
	Unadj.	0.229	0.936	0.649	0.804	2.743	0.157	0.096	1.363
t = B/SE = "Significance" Statistic	Adjusted	0.128	-0.919	1.338	-0.309	2.802	-0.698	-0.163	0.364
	Unadj.	0.371	-0.321	1.444	0.049	2.218	-0.261	-0.229	0.160
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.023	-0.182	0.283	-0.064	0.607	-0.160	-0.047	0.083
	Unadj.	0.097	-0.080	0.385	0.012	0.540	-0.074	-0.067	0.044
N = Number of Schools in Computation	FT	29	29	29	29	29	29	29	29
	NFT	20	20	20	20	20	20	20	20

Figure VII - 3
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 3

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:
Cov. Adj. Effect
Unadj. Effect



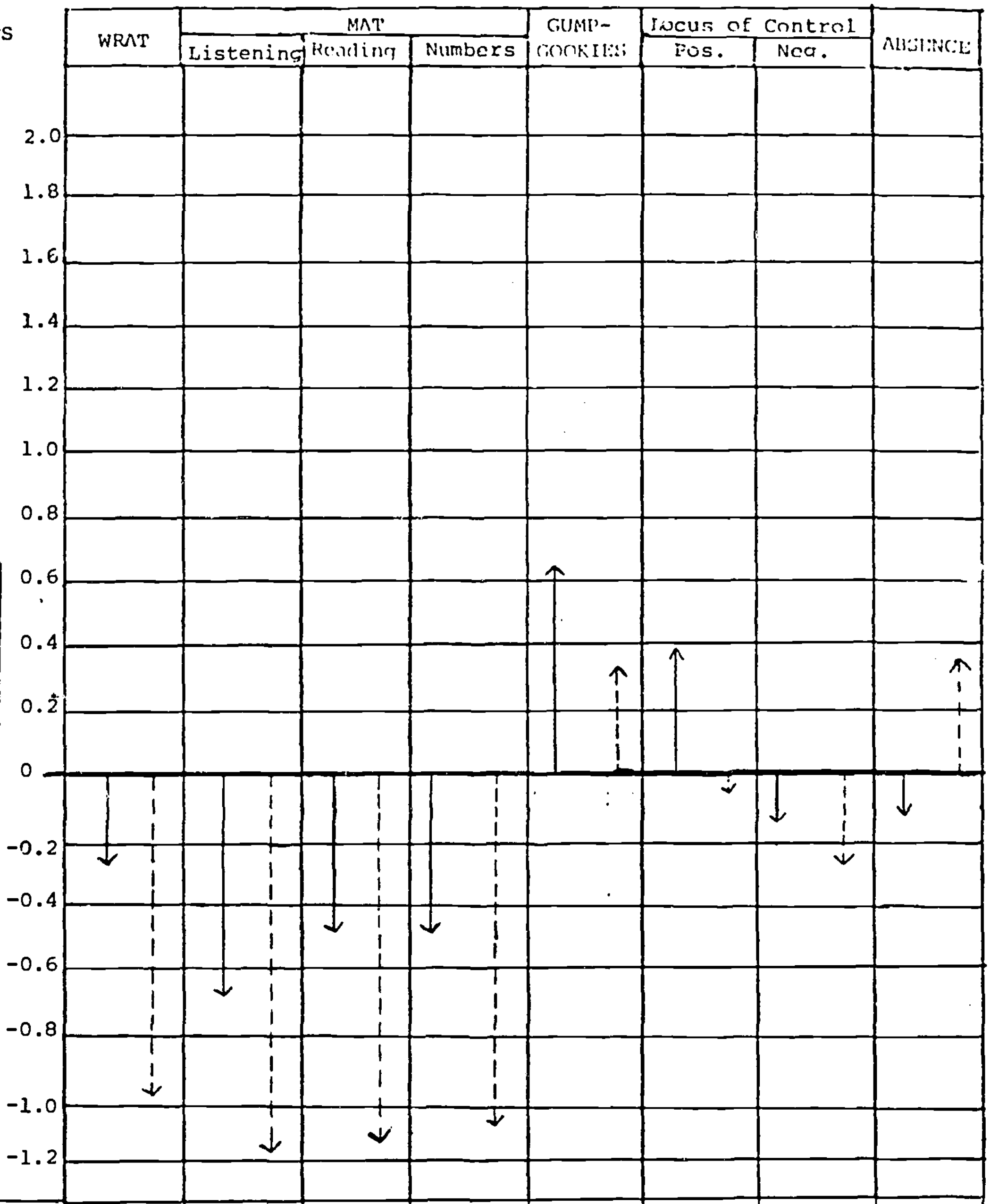
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCES
B = Magnitude of Effect	Adjusted	0.239	-0.833	0.386	-1.463	5.649	-0.145	0.053	-0.754
	Unadj.	-0.149	-1.920	-0.478	-2.488	4.080	-0.351	0.005	1.678
SE = Standard Error of B	Adjusted	0.174	0.817	0.567	0.715	2.689	0.139	0.105	1.252
	Unadj.	0.246	1.006	0.698	0.864	2.949	0.169	0.103	1.465
t = B/SE = "Significance" Statistic	Adjusted	1.374	-1.020	0.681	-2.047	2.101	-1.045	0.504	-0.603
	Unadj.	-0.606	-1.908	-0.685	-2.879	1.384	-2.077	0.049	1.145
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.274	-0.223	0.159	-0.463	0.501	-0.265	0.160	-0.150
	Unadj.	-0.171	-0.513	-0.196	-0.788	0.362	-0.641	0.015	0.335
N = Number of Schools in Computation	FT	21	21	21	21	21	21	21	21
	NFT	20	20	20	20	20	20	20	20

Figure VII - 4
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 5

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:
Cov. Adj. Effect — Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.226	-2.565	-1.201	-1.562	7.118	0.215	-0.040	-0.571
	Unadj.	-0.861	-4.450	-2.769	-3.361	3.860	-0.029	-0.091	1.928
SE = Standard Error of B	Adjusted	0.235	1.103	0.766	0.965	3.629	0.188	0.142	1.690
	Unadj.	0.337	1.379	0.957	1.184	4.041	0.232	0.141	2.001
t = B/SE = "Significance" Statistic	Adjusted	-0.960	-2.326	-1.568	-1.619	1.961	1.147	-0.282	-0.338
	Unadj.	-2.554	-3.227	-2.894	-2.838	0.955	-0.126	-0.641	0.960
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	-0.259	-0.685	-0.493	-0.495	0.631	0.392	-0.120	-0.114
	Unadj.	-0.987	-1.189	-1.138	-1.064	0.342	-0.053	-0.275	0.385
N = Number of Schools in Computation	FT	12	12	12	12	12	12	12	12
	NFT	10	10	10	10	10	10	10	10

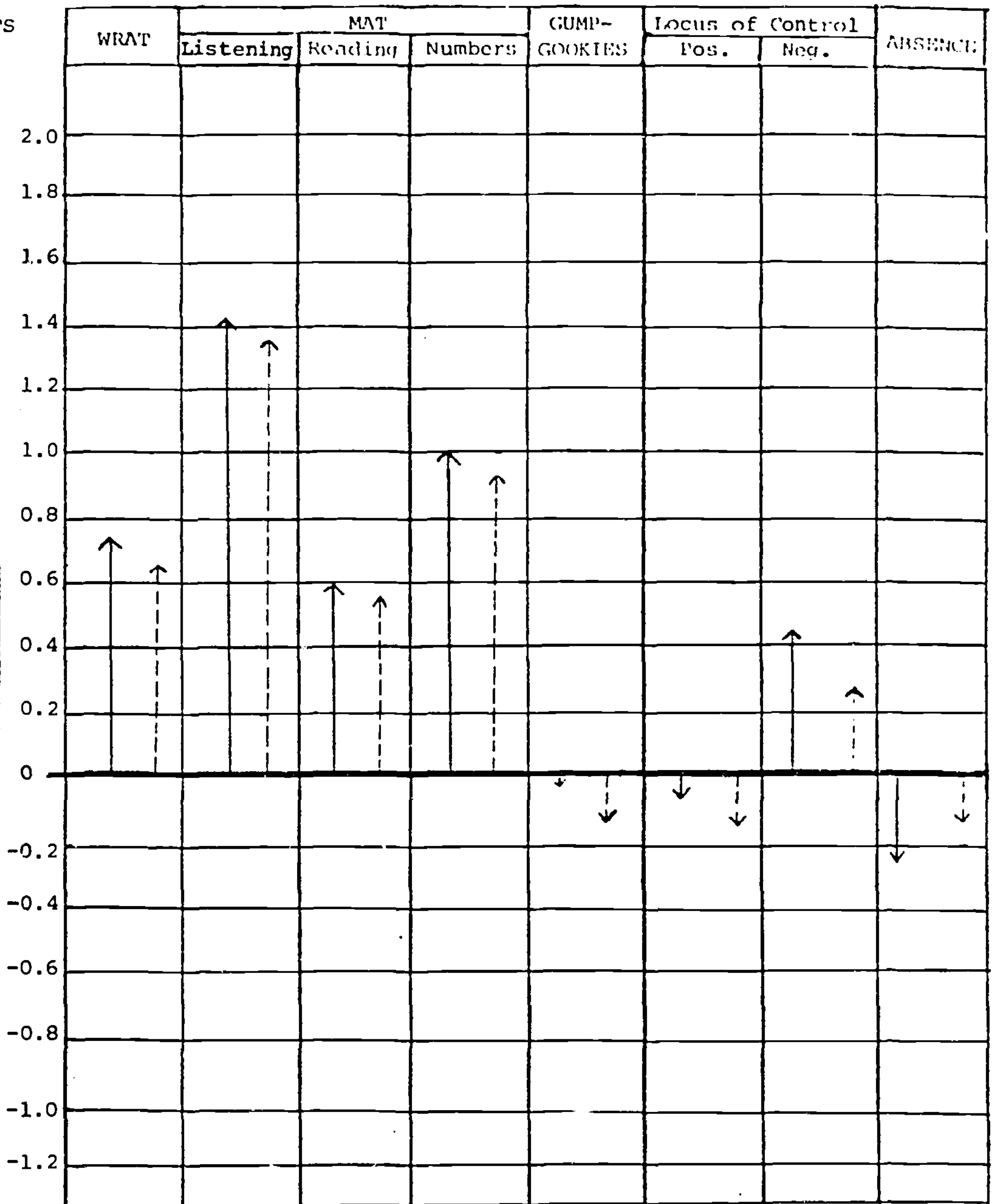
Figure VII - 5
FOLLOW THROUGH EFFECTS PROFILE FOR
SPONSOR 7

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect — Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.659	5.332	1.469	3.154	-0.168	-0.030	0.148	-1.239
	Unadj.	0.564	5.022	1.387	2.923	-1.210	-0.082	0.087	-0.673
SE = Standard Error of B	Adjusted	0.255	1.196	0.830	1.046	3.935	0.204	0.154	1.832
	Unadj.	0.354	1.447	1.004	1.243	4.242	0.243	0.148	2.108
t = B/SE = "Significance" Statistic	Adjusted	2.586	4.459	1.769	3.015	-0.043	-0.148	0.962	-0.676
	Unadj.	1.594	3.470	1.381	2.352	-0.285	-0.336	0.587	-0.319
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.756	1.425	0.604	0.999	-0.015	-0.055	0.448	-0.247
	Unadj.	0.647	1.342	0.570	0.926	-0.107	-0.149	0.264	-0.134
N = Number of Schools in Computation	FT	11	11	11	11	11	11	11	11
	NFT	9	9	9	9	9	9	9	9

Figure VII - 6
FOLLOW THROUGH EFFECTS
PROFILE FOR

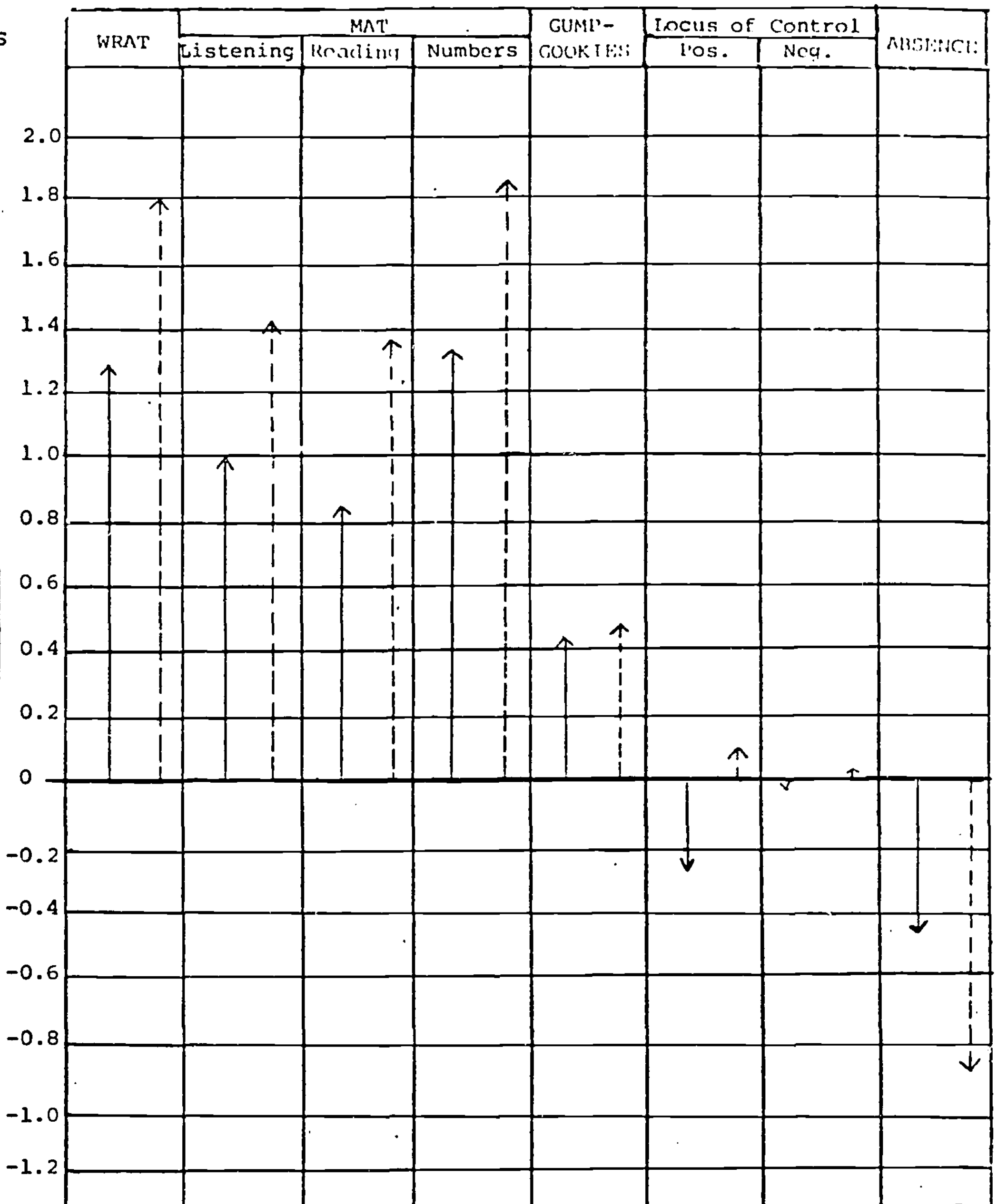
SPONSOR 8

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect — Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	1.124	3.764	2.089	4.378	4.904	-0.151	-0.002	-2.357
	Unadj.	1.571	5.291	3.315	5.781	5.295	0.059	0.004	-4.375
SE = Standard Error of B	Adjusted	0.214	1.002	0.696	0.876	3.297	0.171	0.129	1.535
	Unadj.	0.301	1.230	0.853	1.056	3.604	0.207	0.126	1.791
t = B/SE = "Significance" Statistic	Adjusted	5.252	3.757	3.004	4.997	1.487	-0.887	-0.016	-1.536
	Unadj.	5.219	4.302	3.885	5.475	1.469	0.286	0.032	-2.443
S = Standard Deviation		0.372	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	1.289	1.006	0.858	1.387	0.435	-0.276	-0.006	-0.470
	Unadj.	1.802	1.414	1.362	1.831	0.469	0.108	0.011	-0.873
N = Number of Schools in Computation	FT	16	16	16	16	16	16	16	16
	NFT	12	12	12	12	12	12	12	12

Figure VII - 7
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 9

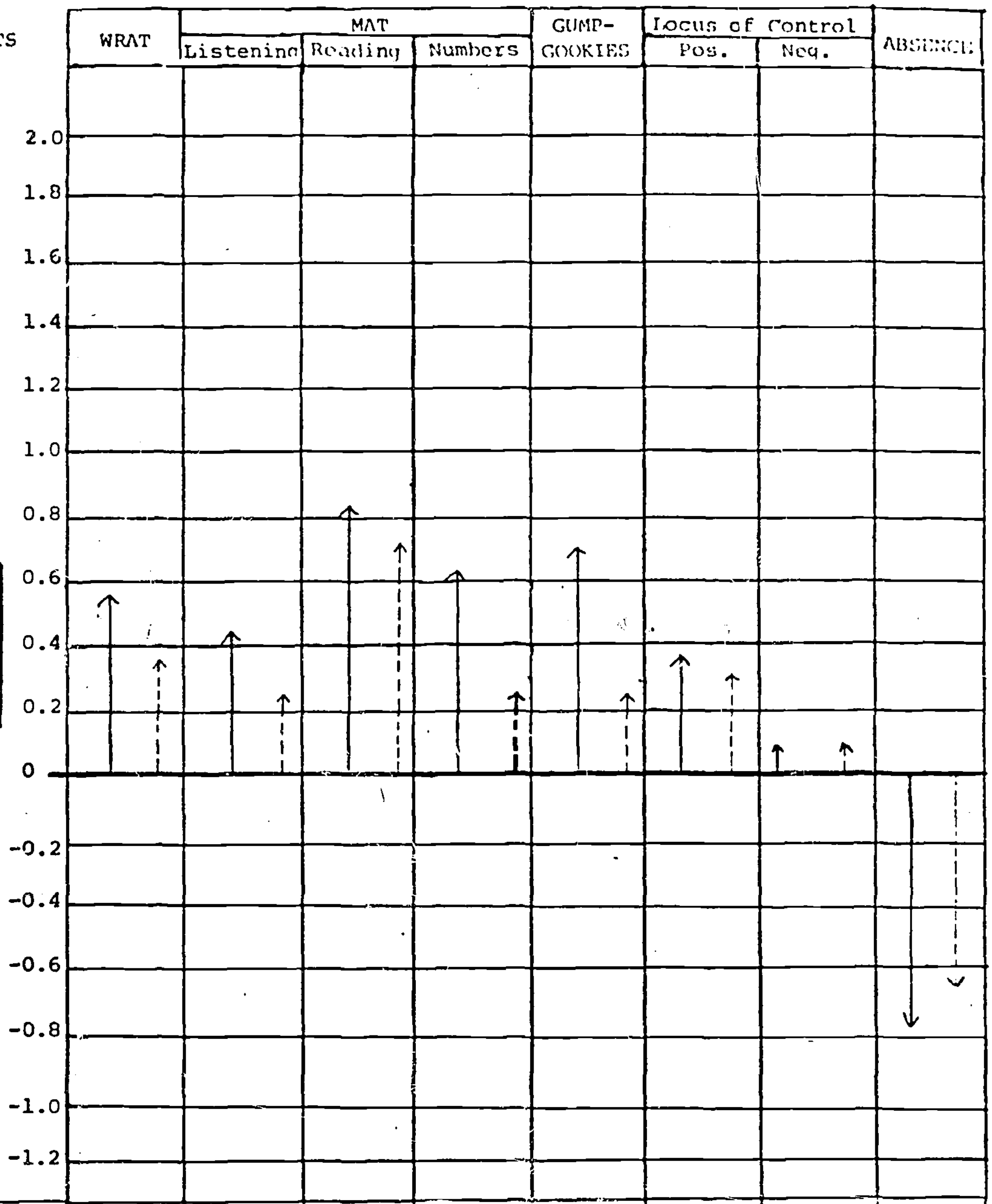
Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect

Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.510	1.696	2.048	2.002	7.489	0.213	0.030	-3.918
	Unadj.	0.298	0.807	1.733	0.873	2.577	0.165	0.031	-3.350
SE = Standard Error of B	Adjusted	0.274	1.284	0.892	1.123	4.226	0.219	0.165	1.967
	Unadj.	0.382	1.565	1.086	1.344	4.586	0.263	0.161	2.278
t = B/SE = "Significance" Statistic	Adjusted	1.862	1.321	2.296	1.782	1.772	0.972	0.184	-1.992
	Unadj.	0.780	0.516	1.597	0.650	0.562	0.629	0.191	-1.470
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.585	0.453	0.842	0.634	0.664	0.387	0.092	-0.781
	Unadj.	0.341	0.216	0.712	0.277	0.228	0.301	0.093	-0.668
N = Number of Schools in Computation	FT	8	8	8	8	8	8	8	8
	NFT	9	9	9	9	9	9	9	9

Figure VII - 8

FOLLOW THROUGH EFFECTS PROFILE FOR

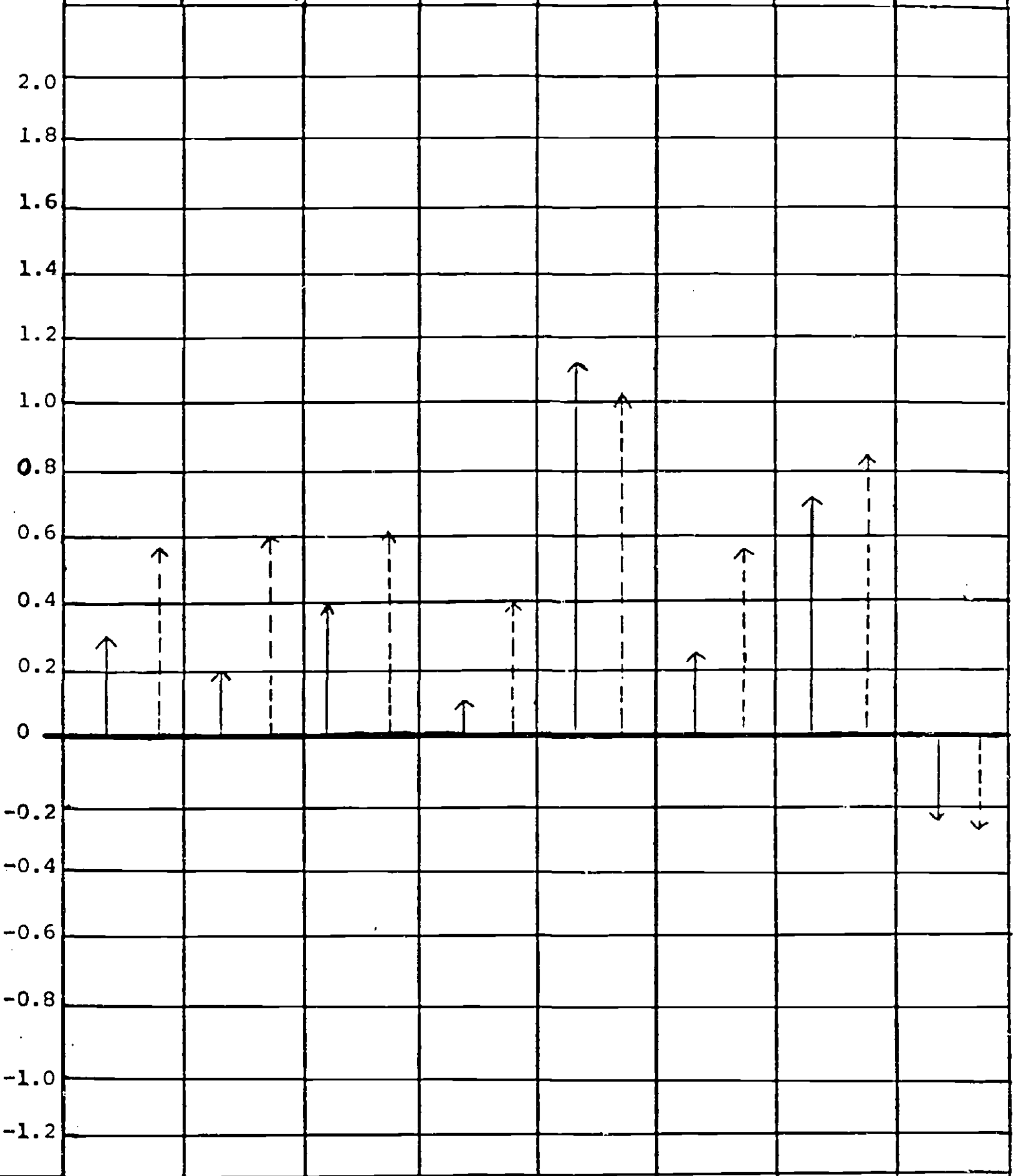
SPONSOR 10

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect   Unadj. Effect

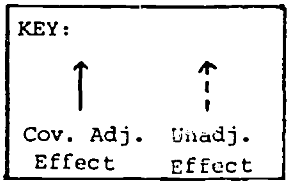


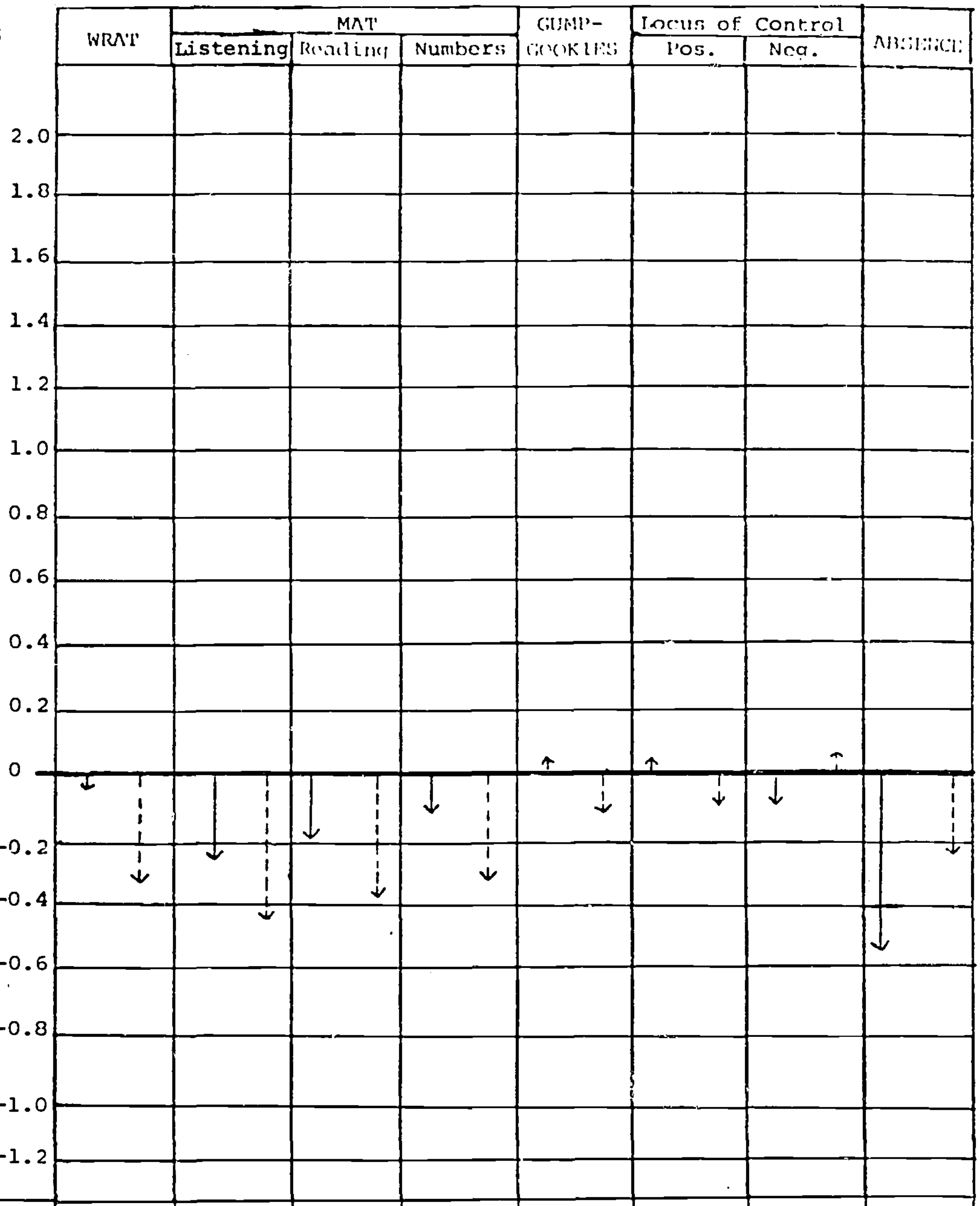
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.275	1.102	0.989	0.286	12.184	0.128	0.255	-1.194
	Unadj.	0.497	2.256	1.491	1.264	11.669	0.315	0.279	-1.291
SE = Standard Error of B	Adjusted	0.240	1.125	0.781	0.984	3.701	0.192	0.145	1.723
	Unadj.	0.332	1.385	0.942	1.166	3.980	0.228	0.139	1.977
t = B/SE = "Significance" Statistic	Adjusted	1.149	0.980	1.266	0.291	3.292	0.667	1.759	-0.693
	Unadj.	1.498	1.628	1.583	1.084	2.932	1.381	2.007	-0.653
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.316	0.194	0.406	0.091	1.080	0.233	0.771	-0.238
	Unadj.	0.570	0.603	0.613	0.400	1.035	0.574	0.844	-0.257
N = Number of Schools in Computation	FT	15	15	15	15	15	15	15	15
	NFT	9	9	9	9	9	9	9	9

Figure VII - 9
FOLLOW THROUGH EFFECTS PROFILE FOR
SPONSOR 11

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



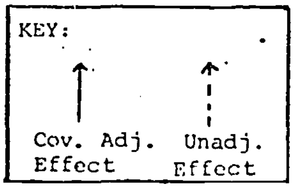


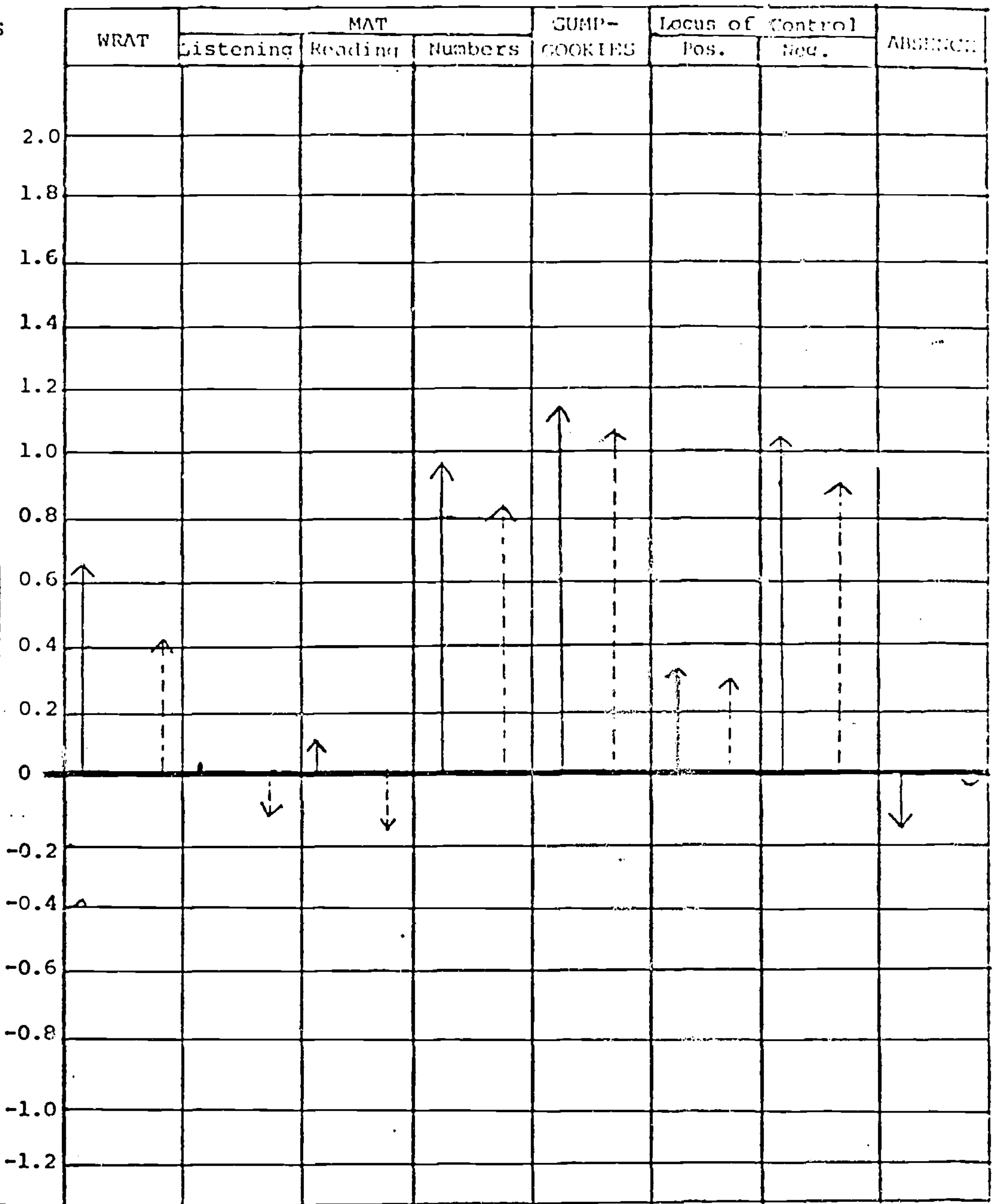
		WRAT	Listening	Reading	Numbers	GUMP-COOKIES	Pos.	Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.027	-0.917	-0.484	-0.369	0.438	0.022	-0.033	-2.859
	Unadj.	-0.290	-1.611	-0.946	-1.032	-1.566	-0.060	0.011	-1.272
SE = Standard Error of B	Adjusted	0.252	1.184	0.822	1.036	3.896	0.202	0.152	1.814
	Unadj.	0.362	1.480	1.027	1.270	4.337	0.249	0.152	2.154
t = B/SE = "Significance" Statistic	Adjusted	-0.106	-0.774	-0.589	-0.357	0.112	0.107	-0.215	-1.576
	Unadj.	-0.801	-1.089	-0.922	-0.813	-0.361	-0.240	0.073	-0.591
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	-0.031	-0.245	-0.199	-0.117	0.039	0.040	-0.099	-0.570
	Unadj.	-0.332	-0.431	-0.389	-0.327	-0.139	-0.108	0.034	-0.254
N = Number of Schools in Computation	FT	10	10	10	10	10	10	10	10
	NFT	9	9	9	9	9	9	9	9

Figure VII - 10
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 12

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)





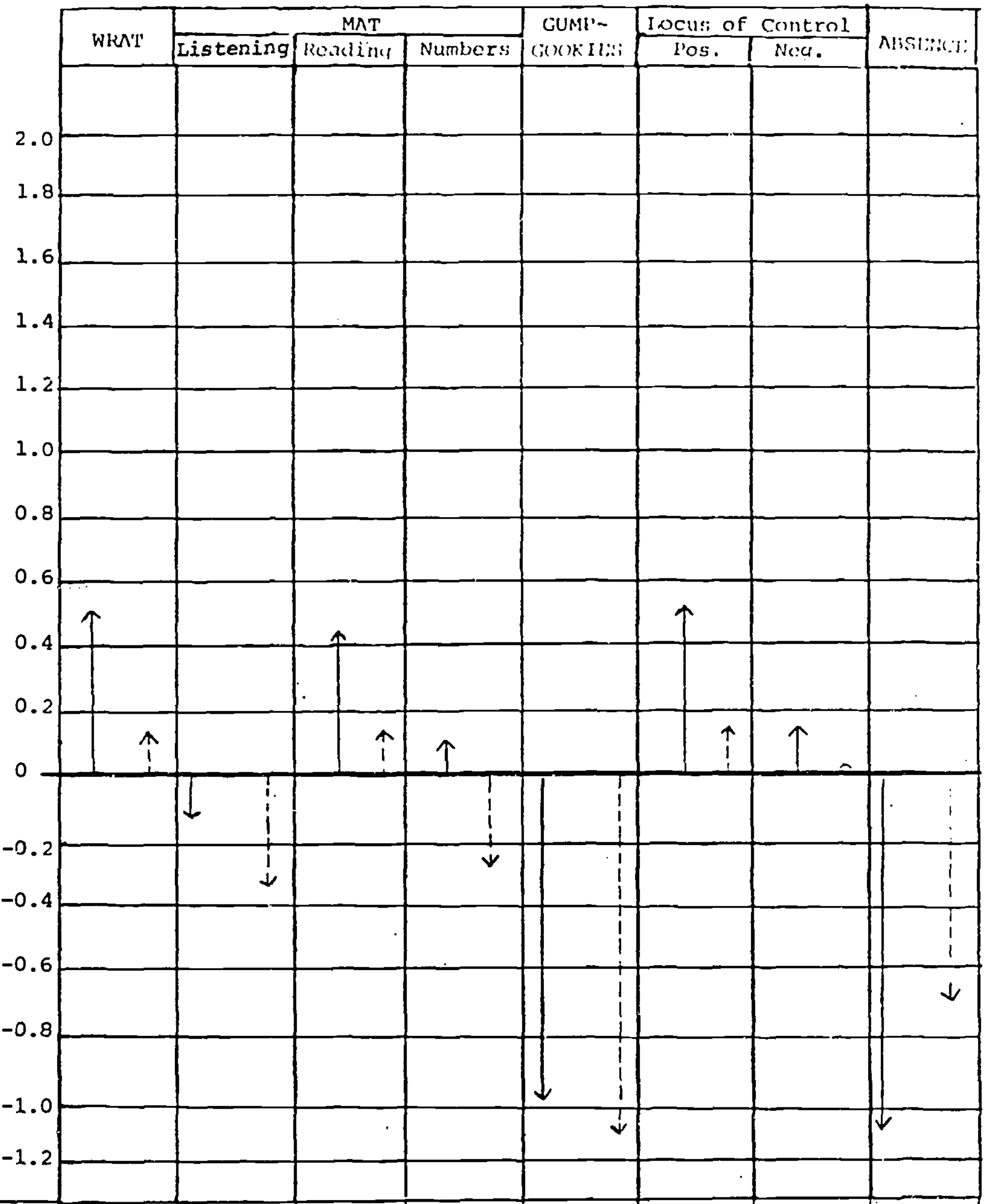
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.562	0.003	0.274	3.059	12.991	0.172	0.341	-0.7[illegible]
	Unadj.	0.359	-0.396	-0.346	2.611	12.039	0.170	0.302	-0.1[illegible]
SE = Standard Error of B	Adjusted	0.253	1.187	0.824	1.038	3.907	0.202	0.153	1.[illegible]
	Unadj.	0.354	1.447	1.004	1.243	4.242	0.243	0.149	2.1[illegible]
t = B/SE = "Significance" Statistic	Adjusted	2.222	0.003	0.332	2.946	3.325	0.852	2.229	-0.4[illegible]
	Unadj.	1.015	-0.274	-0.345	2.101	2.838	0.700	2.027	-0.[illegible]
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.[illegible]
*B/S = Effect in Standard Deviations	Adjusted	0.645	0.0008	0.113	0.969	1.152	0.314	1.033	-0.1[illegible]
	Unadj.	0.412	-0.106	-0.142	0.827	1.067	0.309	0.913	-0.[illegible]
N = Number of Schools in Computation	FT	9	9	9	9	9	9	9	9
	NFT	11	11	11	11	11	11	11	11

Figure VII - 11
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 14

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY: Cov. Adj. Effect; Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.445	-0.466	1.046	0.325	-10.886	0.306	0.050	-5.263
	Unadj.	0.130	-1.285	0.382	-0.900	-12.398	0.074	0.001	-3.465
SE = Standard Error of B	Adjusted	0.331	1.551	1.077	1.357	5.104	0.264	0.200	2.376
	Unadj.	0.477	1.950	1.353	1.674	5.715	0.327	0.200	2.839
t = B/SE = "Significance" Statistic	Adjusted	1.346	-0.300	0.971	0.239	-2.133	1.157	0.252	-2.215
	Unadj.	0.273	-0.659	0.283	-0.537	-2.169	0.225	0.006	-1.220
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.510	-0.124	0.430	0.103	-0.965	0.557	0.152	-1.050
	Unadj.	0.149	-0.343	0.157	-0.285	-1.099	0.134	0.004	-0.691
N = Number of Schools in Computation	FT	6	6	6	6	6	6	6	6
	NFT	5	5	5	5	5	5	5	5

(EDC) shows no effect that we can have any confidence in.

The data on absences deserve particular attention here. This variable can be influenced by a variety of factors which cannot yet be untangled. A child may have fewer absences than another because (1) he finds school a more attractive place and wishes to attend, (2) his parents feel that school is more attractive and urges the child to attend, (3) the child is sick less often than others, (4) any combination of the above. Although none of these factors can be separated from the others with the data available, all suggest positive characteristics of a program associated with fewer absences. In the case of Follow Through, with mandated social and health services, fewer absences due to fewer sicknesses represents a positive program effect.

In the aggregate, FT children tended to be absent 1.9 fewer days than their NFT comparisons. There is considerable Sponsor variability on this outcome: the children attending Sponsor 14 (SEDL) schools were absent five days less than the children attending the NFT schools; Sponsor 9 (High Scope) children were absent 3.9 fewer days; Sponsor 11 (EDC) children were absent 2.4 fewer days. The remaining Sponsors all had children who were absent approximately the same number of days as the children attending the NFT schools. In no case were the FT children absent more often than the NFT children. Once again, the basis of this FT advantage in attendance cannot be established, but it is clear that for whatever the reasons, given the number of children attending the FT programs throughout the nation, these findings indicate significant impact on the total average daily attendance in the participating school districts.

1.3.3 Effects of Covariance Adjustment

NFT schools were selected judgmentally to be comparable to the FT schools that had already been chosen. Since our estimates of FT's success depend critically on these comparison schools, it is important to know how well FT and NFT are matched, both overall and Sponsor by Sponsor.

Table VII-4 casts some indirect but informative light on this question. The tabulated numbers are the algebraic differences between the adjusted and unadjusted FT effects, expressed in criterion standard deviations. The table tells us, for example, that covariance adjustment increased Sponsor 3's (University of Arizona's) WRAT effect by .445 standard deviations while decreasing the same Sponsor's Absence effect (making it more negative and therefore larger) by .485 standard deviations. All of Arizona's adjustment effects are positive except the Absence adjustment, suggesting that the adjustment compensated for a generalized initial disadvantage of these FT schools with respect to their NFT comparison schools: for Sponsor 3, FT started out substantially "behind" NFT academically and this disadvantage extends across all eight of our outcome measures. A similar pattern holds for Sponsors 5, 7, 12, and 14 and, with minor deviations, for Sponsors 9 and 11. FT seems to have started out behind NFT at various degrees in these seven Sponsors. The pattern is reversed for Sponsors 8 and 10: here, FT started out ahead of NFT. Only in Sponsor 2 (Far West) is the match close enough to make direct comparisons unequivocal.

The "mismatch index" in the last column of Table VII-4 summarizes this "behindness" across the eight criterion variables: it is the mean of each Sponsor's eight adjustment effects, with the sign of the Absence effect reversed. It suggests that Sponsors 3, 5, 8, and 14 contain relatively severe FT/NFT mismatches and that we should therefore be especially careful in interpreting the adjusted effects for these Sponsors: where the adjustment procedure must work hardest, we must be most aware of the possible consequences of its shortcomings.

It should be noted that there may be idiosyncratic bases for

TABLE VII - 4

Effects of Covariance Adjustment
in Criterion Standard Deviation Units

SPONSOR	WRAT	M A T			GUMP	Locus of Control		ABSENCE	MISMATCH INDEX*
		WORD	READ	NUM		+	-		
2	-.074	-.102	-.102	-.076	.067	-.086	.020	-.039	-.039
3	.445	.290	.355	.325	.139	.376	.145	-.485	.322
5	.728	.504	.645	.569	.289	.445	.155	-.499	.417
7	.089	.083	.034	.073	.092	.094	.084	-.113	.083
8	-.513	-.408	-.504	-.444	-.034	-.384	-.017	.403	-.339
9	.244	.237	.130	.357	.336	.086	-.001	-.113	.188
10	-.254	-.409	-.207	-.309	.045	-.341	-.073	.019	-.196
11	.301	.186	.190	.210	.278	.147	-.133	-.316	.187
12	.233	.107	.255	.142	.085	.005	.020	-.139	.123
14	.361	.219	.273	.388	.134	.423	.148	-.359	.276
MAIN EFFECT	.158	.081	.107	.123	.143	.077	.055	-.157	.113

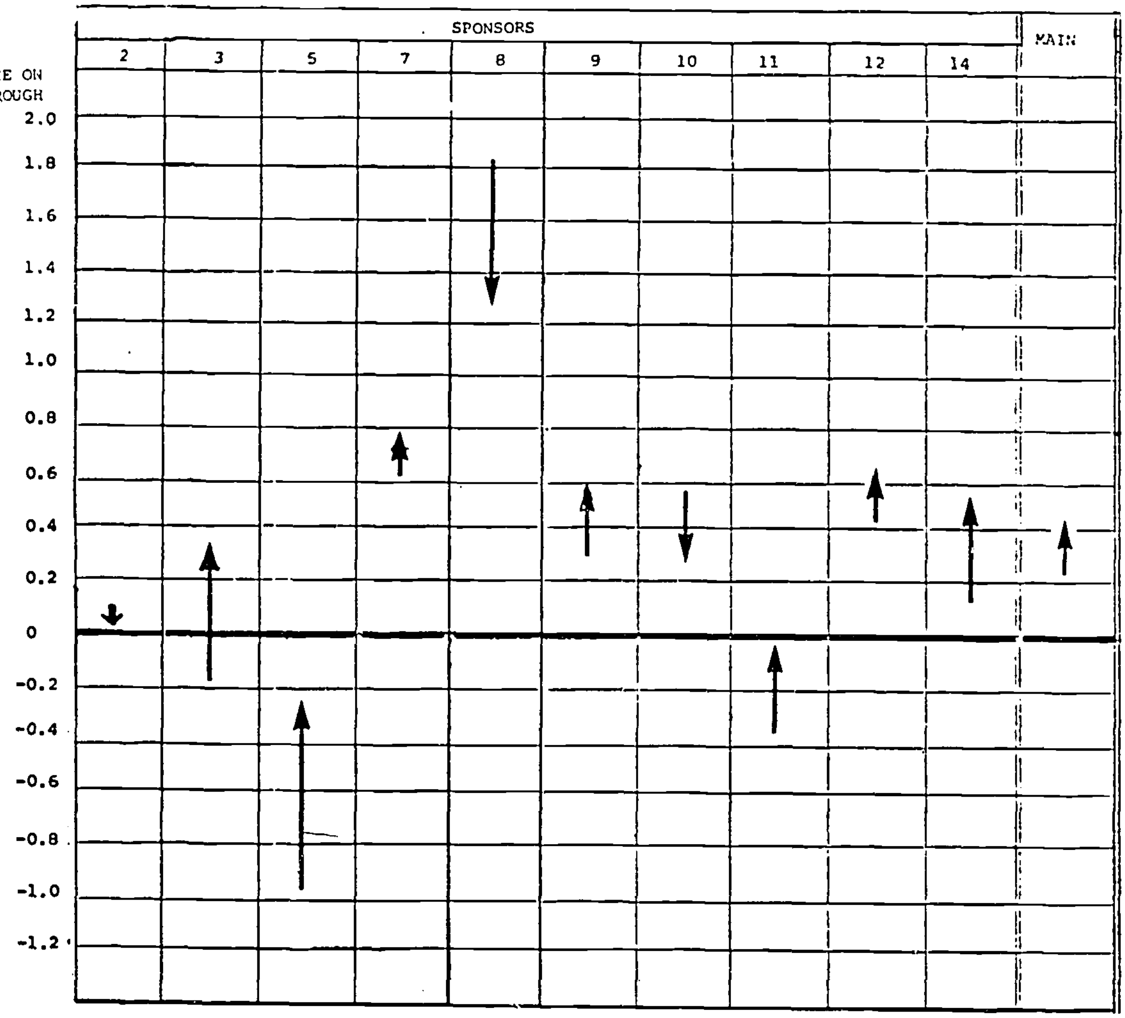
*The mean of the eight adjustment effects, with the sign of the Absence effect reversed.

these mismatches. For example, Sponsor 8 (University of Kansas) has indicated that a great deal of instruction occurs in the very first weeks of school, whereas Fall testing did not usually occur until four or more weeks had elapsed. Thus the measured superiority of the FT classes may actually be Sponsor effect. This possibility is discussed in section 1.4 of this chapter.

A final graphical display of the mismatch data raises one further issue. Figure VII-12 gives us at least an heuristic handle on the possible role of fan spread in these effects. (The problem of fan spread was introduced in Chapter II.) According to this hypothesis, those who start out ahead in the achievement race got there by achieving more rapidly before treatment started. In the absence of any effective treatment, the gap continues to widen by sheer inertia. In the case, then, where FT starts out behind NFT, FT must make up not only the initial deficit but also the additional disadvantage generated over time by the difference in rates of progress. Fan spread thus militates against the detection of real FT effects when FT starts out behind NFT. By the same token, fan spread enhances effects spuriously when the initial mismatch is in the opposite direction. A "pure fan spread pattern" (in Figure VII-12) would have all arrows pointing toward the zero-axis, with shorter arrows close to the axis and longer ones farther from it. In fact, four of the ten Sponsors' arrows point toward the axis, four point away from it, and the directions of two are equivocal. It is true that the two longest arrows are also associated with the two largest Sponsor effects, and that these both point inward, in accordance with the fan spread hypothesis. One might therefore plausibly suspect that Sponsor 5's (Bank Street's) effects were obscured by fan spread and that Sponsor 8's (University of Kansas') effects were spuriously inflated by fan spread. The pattern of the other Sponsors' adjustments, however, is not all consistent with the fan spread hypothesis. While we do not yet have the longitudinal data that we would need in order to adjust for fan spread, we take this pattern as evidence that the apparent effects of FT are not merely fan spread artifacts: something else is happening.

Figure VII - 12

EFFECT OF COVARIANCE ON THE WRAT FOLLOW THROUGH EFFECTS OF TEN SPONSORS AND THEIR AGGREGATE



1.4.0 THE EFFECT OF PRETEST DELAY[3]

The problem of FT/NFT mismatch is a central issue in the evaluation of Project Follow Through. Section 1.3.3 presents the hypothesis that FT/NFT differences on the Fall WRAT reflect early treatment effects. There is some information available at the school level of analysis to investigate this assertion. This information is the number of days from the beginning of school to the date of pretest, called pretest delay. If treatment effects are occurring in the first few weeks of school for a given Sponsor, it is possible that the first schools pretested in the Fall will obtain lower WRAT scores than those pretested later in the Fall. If this is the case, then a simple correlation of pretest delay with Fall WRAT will be positive and significantly different from zero.

It is important to point out that neither a rejection nor an acceptance of the "early effects" hypothesis is possible on the basis of these zero-order correlations. First, the values of the variable, pretest delay,cover a limited number of days; hence the lack of a positive linear relationship (as measured by the simple correlation) between these values and the corresponding Fall WRAT scores cannot eliminate the possibility of a general rise in the Fall scores over the early weeks of kindergarten due primarily to treatment effects. Second, the time of testing study has produced some evidence of a non-random testing schedule; hence any significant correlation, or lack thereof, might be an artifact of a biased schedule (i.e., who was tested when?) rather than a reflection of early treatment effects.

Table VII-5 presents the correlations between pretest delay and pretest scores, as well as the corresponding two-tailed probability levels, for each Sponsor by FT/NFT group.

As might be expected because of the small range of values for the delay variable, many Sponsors produce correlations which are not significantly different from zero. Of the five Sponsors (2, 3, 5, 7, and 9) who do have significant correlations for their FT groups, three Sponsors (2, 3, and 7) have similar correlations for their NFT groups.

[3]The information presented here is part of a study on the effect of the testing schedules on the data utilized in this report.

TABLE VII - 5

Zero-Order Correlations
Of Pretest Delay with Fall WRAT Scores
By Sponsor by FT/NFT

	FT		NFT	
Sponsor	Correlation	p*	Correlation	p*
2	-.4868	.01	-.6388	.00
3	-.3410	.10	-.3434	.14
5	.4735	.06	-.2317	.42
7	.6335	.04	.5994	.07
8	-.1245	.61	-.0596	.83
9	-.6206	.08	-.3527	.32
10	.2406	.37	-.4634	.21
11	.2282	.45	.2272	.48
12	-.2213	.57	.0272	.94
14	-.2127	.58	.0590	.89

*Probability levels are based on two tailed significance tests.

Sponsor 9 (High/Scope) has a large negative correlation between pretest delay and Fall WRAT scores, i.e., the schools tested later in the Fall tend to have lower pretest scores than those tested earlier. Sponsor 5 (Bank Street) it the only Sponsor whose correlations suggest a possible early treatment effect. The FT schools with a longer delay between the first day of school and date of pretest tend to have higher Fall WRAT scores than the other FT schools.

Also of interest is the Kansas program (Sponsor 8). Table VII-5 reflects a non-significant correlation for Sponsor 8's FT and NFT groups. Hence, within the range of pretest administration there was no difference in pretest scores for this Sponsor's FT schools. Although this does not support the early treatment effect hypothesis as explained earlier, we cannot reject this hypothesis at this time.

The data suggest that the FT advantage in Fall WRAT for this Sponsor might reflect the skills with which the FT children entered the program, as much as a treatment effect. This issue must be examined further before the large posttest advantages can be ascribed to either fan spread or treatment effects.

1.5.0 DISCUSSION

Let us note three salient generalizations pertaining to this one year study of kindergarten effects:

1. The variability of mismatch of FT and NFT groups among Sponsors is quite large. This suggests that the local conditions faced by Sponsors, which resulted in the local assignment of schools to FT or NFT status, varied considerably. It is very likely that this variability occurred among sites within each Sponsor. These local conditions are critical to the understanding of the educational experiences delivered by each Sponsor at each site. The present analyses must be considered incomplete until these factors are assessed and entered into the analyses.

2. Variability in the pattern of outcomes among Sponsors is great enough to preclude grouping Sponsors into clusters, at this time. At the kindergarten level, Sponsors show every conceivable pattern across achievement, motivations and absence measures, and no two Sponsors show the same pattern. Longitudinal data are required before these patterns can emerge as stable enough to relate to educational inputs.

3. Despite the variability of Sponsor effects, an accumulation of effects across all Sponsors reveals consistent overall FT effects on all measures.

2.0 BIG CITY STUDY

2.1.0 INTRODUCTION

One characteristic of the Follow Through analytic sample is its uneven distribution across Sponsors. In order to obtain more evenly matched subsamples, and to minimize regional and population density variation, several Sponsors were concentrated in three big cities: New York, Philadelphia, and Chicago. The purpose of this study is to look at the effects of FT in the sample schools in these cities relative to effects in lower density areas. More specifically, the primary questions we want to investigate are:

- What is the FT main effect in Big City schools relative to non Big City schools?
- What are the individual Sponsor effects in Big City schools relative to non Big City schools?

A limitation of this study is that there are only 37 Big City schools in the sample. Because of this small sample size and the large number of variables involved in this study, we cannot examine Big City schools directly.[4] Consequently, we will indirectly investigate the effect of Big City schools by comparing the results of the 251 non Big City schools reported earlier in this chapter with those of the total sample of 288 schools, which we analyze here.

2.2.0 METHOD

2.2.1 Analytic Subset

Test scores used in this study were obtained from the Cohort III kindergarten sample. The combined (Big Cities and non Big Cities) sample, which will be analyzed in this study, consists of 288 schools distributed across 10 Sponsors (see Table VII-6). Of these, 37 schools representing seven Sponsors constitute the Big Cities sample. Table VII-7 gives this distribution. Sponsors 2, 3, and 12 are not involved individually in this study since they have no Big City schools; they are, however, included in the FT/NFT main effect results.

---

[4] In order to estimate effects, the sample size must be considerably larger than the number of variables. This is because each sample unit provides a degree of freedom for estimating effects, while each variable uses up one degree of freedom.

TABLE VII - 6

Distribution of the Complete Analytic Subset
of FT and NFT Schools Across Ten Sponsors

SPONSORS (by code number)

	2	3	5	7	8	9	10	11	12	14	TOTALS
FT	29	21	16	11	20	11	17	13	9	9	156
NFT	20	20	15	10	15	12	9	12	11	8	132
TOTALS	49	41	31	21	35	23	26	25	20	17	288

TABLE VII - 7

Distribution of Big City Schools Across Sponsors' FT and NFT Populations

		SPONSORS										
		2	3	5	7	8	9	10	11	12	14	TOTALS
FT	N	0	0	4	0	4	3	2	3	0	3	19
	%	0%	0%	25%	0%	20%	27%	12%	33%	0%	33%	12%
NFT	N	0	0	5	1	3	3	0	3	0	3	18
	%	0%	0%	33%	10%	20%	25%	0%	27%	0%	28%	14%
TOTALS	N	0	0	9	1	7	6	2	6	0	6	37
	%	0%	0%	29%	5%	20%	26%	8%	30%	0%	35%	13%

Tabulated percentages refer to cell totals for the entire population.

2.2.2 Analytic Design

In order to compare the results of the total sample with those of the non Big Cities sample, the same analysis design is used here, viz. analysis of covariance. The major function of this design is to identify the Sponsor and FT main effects on the criterion variables, adjusting for initial differences both among Sponsors and between FT/NFT groups. Thus, in our analysis of covariance design, Sponsors and FT/NFT groups are examined as independent variables; sets of covariables are utilized as adjusting variables for Sponsor mismatches; and eight criterion measures are analyzed separately.

2.2.3 Variables

Twenty-one variables are included in these school level analyses: eight criterion or outcome measures (WRAT; MAT: Listening to Sounds; Reading, Arithmetic; Gumpgookies; Locus of Control: positive and negative; and Absence); two indicators of "treatment" (FT and NFT), and eleven covariables (Fall WRAT; percentage of Black pupils; percentage of minority pupils; years at current address; adjusted income level; mother's education; parent-school receptivity; western region; southern region; metropolitan area; and middle-sized cities), aggregated to school level. Each is listed and explained in Section 1.2.3.

2.3.0 RESULTS

As mentioned previously we investigate the effect of the Big Cities by comparing the results of the sample including Big City schools with the results of the sample excluding them. Thus, Figures VII-13 through VII-20 on the following pages must be compared with the parallel Figures VII-1, 4, 5, 6, 7, 8, 9, and 11.

For an explanation of the statistics contained in the figures refer to Chapter V.

2.3.1 Main Effects

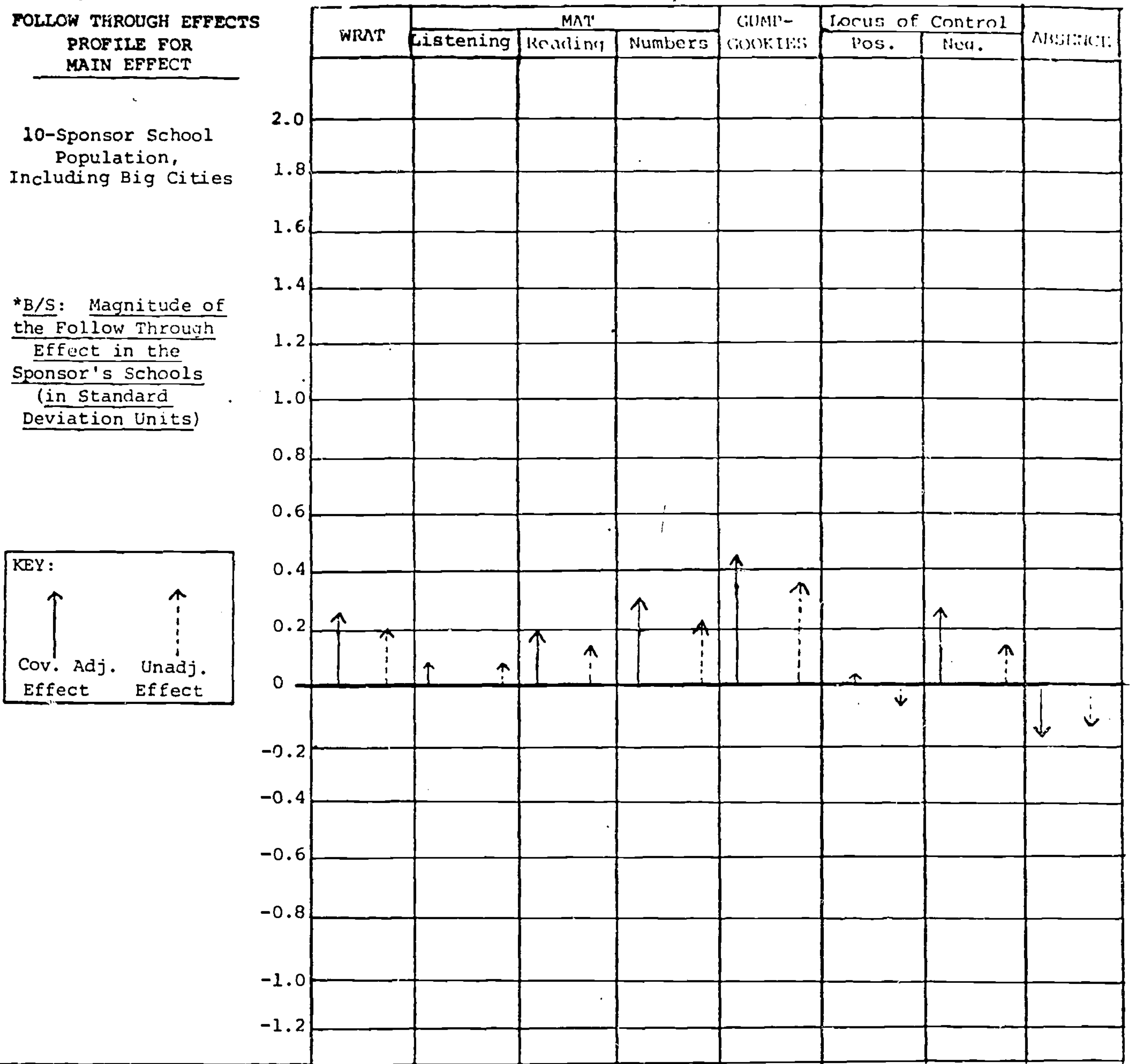
By comparing Figure VII-13 here with Figure VII-1, we can see that the main effects are generally smaller, and covariate adjustment changes

Figure VII - 13

FOLLOW THROUGH EFFECTS PROFILE FOR MAIN EFFECT

10-Sponsor School Population, Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.236	0.312	0.488	0.832	5.512	0.022	0.093	-0.983
	Unadj.	0.177	0.308	0.400	0.669	4.391	-0.032	0.046	-0.843
SE = Standard Error of B	Adjusted	0.082	0.386	0.262	0.319	1.463	0.066	0.051	0.612
	Unadj.	0.098	0.401	0.275	0.335	1.456	0.069	0.044	0.678
t = B/SE = "Significance" Statistic	Adjusted	2.878	0.808	1.864	2.608	3.768	0.333	1.824	-1.606
	Unadj.	1.803	0.767	1.455	2.000	3.016	-0.466	1.045	-1.243
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.267	0.084	0.200	0.268	0.440	0.039	0.263	-0.171
	Unadj.	0.201	0.083	0.164	0.216	0.350	-0.057	0.131	-0.146
N = Number of Schools in Computation	FT	156	156	156	156	156	156	156	156
	NFT	132	132	132	132	132	132	132	132

them less, with Big Cities included. This suggests that FT and NFT schools in the Big Cities are more alike both before and after "treatment" than is the case in the population which excludes the Big Cities. As one might expect, matching was easier in the Big Cities.

2.3.2 Sponsor Effects

When we display the criterion profiles for each of the Sponsors who have Big City schools (Sponsors 5, 7, 8, 9, 10, 11, 14), we see the same general trend here (Figures VII-14 through VII-20) as was observed with Big Cities included (Figures VII-4, 5, 6, 7, 8, 9, and 11); however, Big City schools do cause some differences:

- Sponsor 5 (Bank Street College) shows achievement effects which are smaller (and, since they are negative, therefore "better") with Big Cities included. Here, as in the main effect, the matching picture improved. When the Big Cities are included this Sponsor's marginal positive effect on Locus of Control vanishes, however.

- The same achievement pattern appears in Sponsor 8 (University of Kansas), but much less marked. Achievement motivation increases slightly, and the former Absence effect washes out entirely, with the Big Cities included.

- Sponsor 9 (High/Scope Foundation) has FT effects which are still positive but diminished somewhat (with the exception of Locus of Control) when Big City schools are added.

- In Sponsor 10 (University of Florida), WRAT and achievement motivation effects are decreased, but the MAT Reading effect is increased, with inclusion of Big City schools.

- Sponsor 11 (Educational Development Center), shows a sizable increase in achievement motivation effect when Big City schools are added, but achievement effects become slightly more negative.

- The Big City schools eliminate Sponsor 14's (Southwest Educational Development Laboratory) positive WRAT and MAT Reading effects. The effects pattern here is relatively unstable, however, because of the small number of schools involved.

2.4.0 DISCUSSION

The preceding results would seem to imply that FT is, in general, having somewhat less effect in the Big City schools than in other areas.

Whether this difference is significant or not is difficult to judge because of the necessarily indirect method used in the analysis. The decreased effects which do exist may be due to the greater difficulty Sponsors found in implementing their models in Big City schools. The size and bureaucracy of these school systems, along with a crowded environment which affects teacher, parent, and child attitudes, can contribute to more difficult implementation and thus decrease the effects of Follow Through in the Big City schools. On the other hand, the task of changing the performance of children in large metropolitan centers may be considerably more difficult than the accomplishment of this task elsewhere in the country. The problem may hinge at least as much on the fact of highly bureaucratized school systems, crowded conditions, and political conflicts between community and school, as it does on the failure to implement innovative programs. Under any conditions, it is necessary to measure the actual degree of program implementation within and without the Big Cities before this issue can be settled. The finding that the consequences of FT seem to be different depending upon the local site conditions is not surprising and certainly one which needs to be explored in future studies.

Figure VII - 14

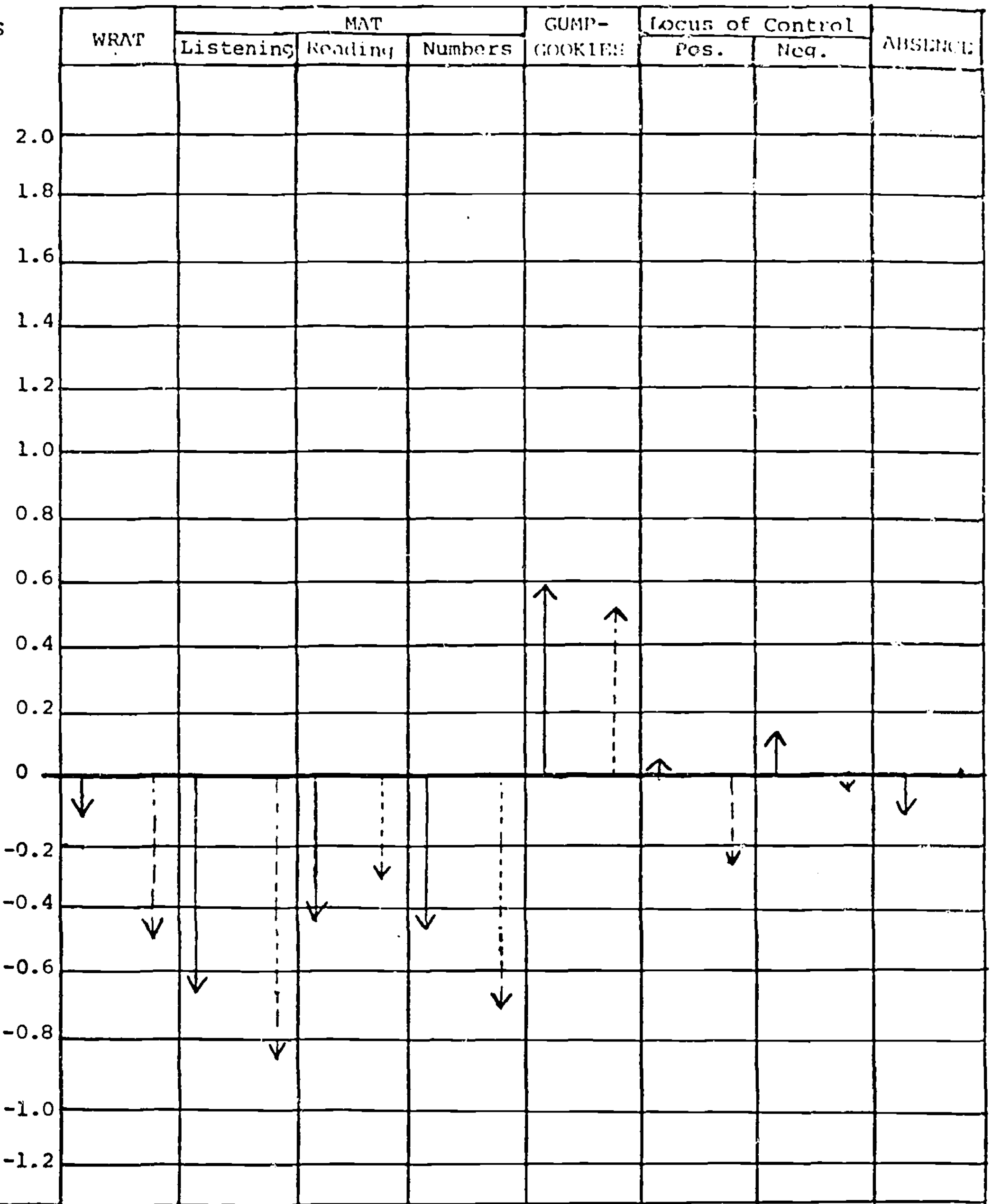
FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 5

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect — Unadj. Effect



		WRAT	Listening	Reading	Numbers	GUMP-COOKIES	Pos.	Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.094	-2.491	-1.022	-1.396	7.495	0.019	0.058	-0.899
	Unadj.	-0.452	-3.187	-1.839	-2.172	6.396	-0.139	-0.011	0.042
SE = Standard Error of B	Adjusted	0.204	0.959	0.651	0.793	3.639	0.164	0.127	1.696
	Unadj.	0.283	1.155	0.791	0.962	4.188	0.200	0.127	1.951
t = B/SE = "Significance" Statistic	Adjusted	-0.462	-2.597	-1.571	-1.760	2.060	0.118	0.459	-0.530
	Unadj.	-1.598	-2.759	-2.325	-2.258	1.527	-0.697	-0.085	0.021
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	-0.107	-0.669	-0.420	-0.450	0.598	0.034	0.165	-0.156
	Unadj.	-0.512	-0.856	-0.310	-0.700	0.510	-0.245	-0.031	0.007
N = Number of Schools in Computation	FT	16	16	16	16	16	16	16	16
	NFT	15	15	15	15	15	15	15	15

Figure VII - 15

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 7

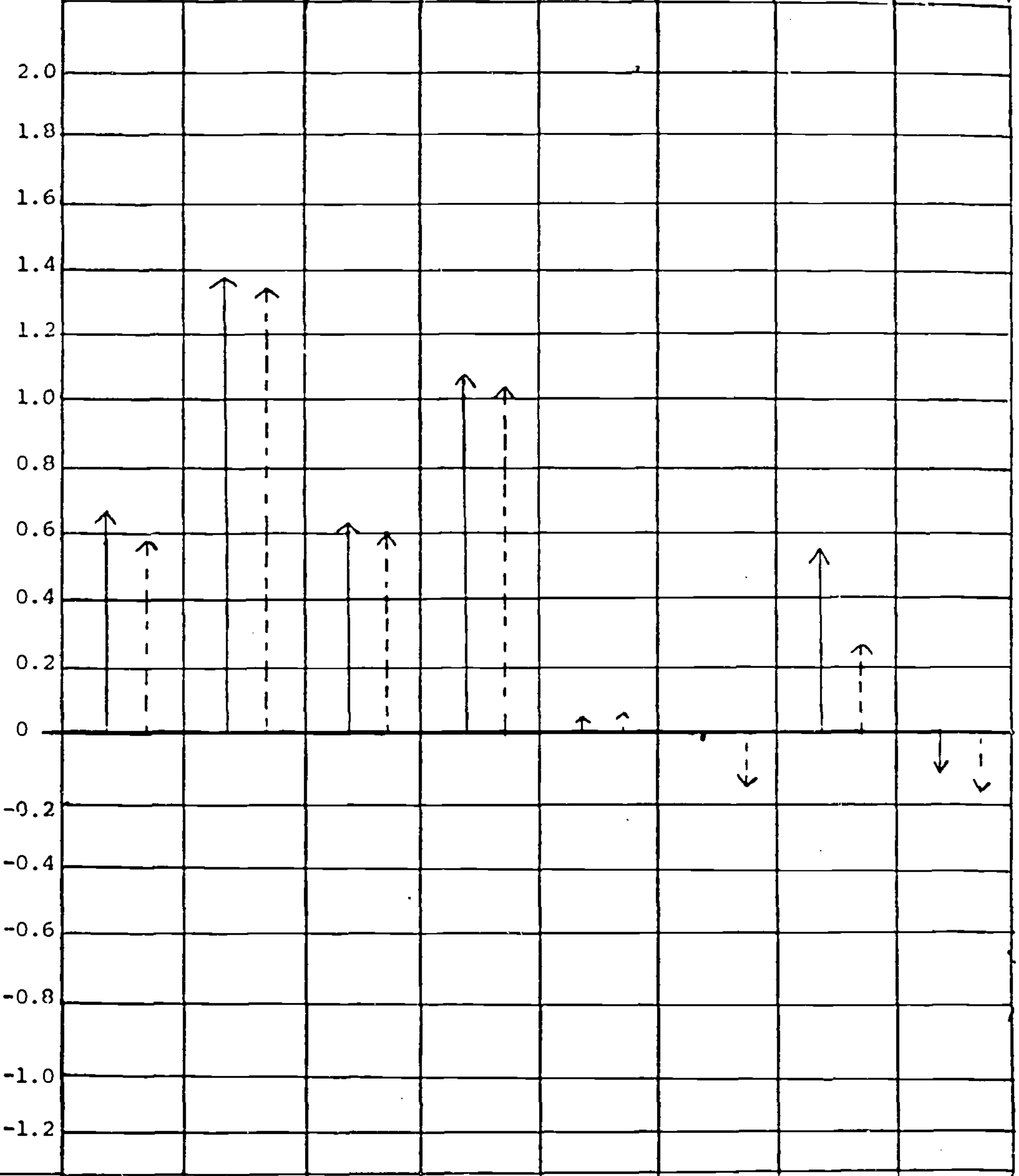
Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect

Unadj. Effect



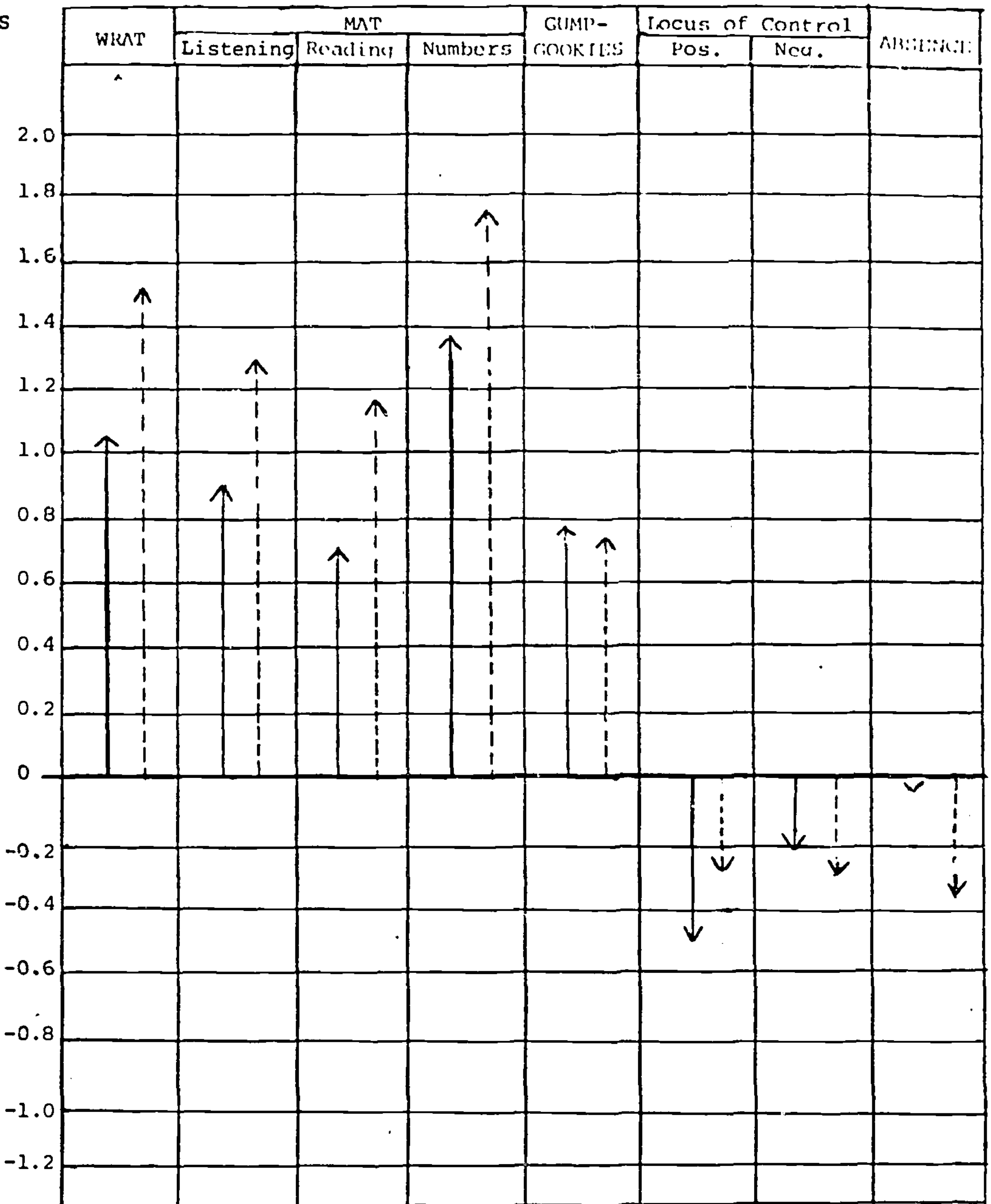
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.595	5.202	1.503	3.308	0.560	-0.0003	0.195	-0.819
	Unadj.	0.528	5.114	1.460	3.199	0.621	-0.092	0.097	-0.978
SE = Standard Error of B	Adjusted	0.251	1.181	0.801	0.976	4.479	0.202	0.156	2.087
	Unadj.	0.344	1.404	0.962	1.170	5.091	0.243	0.154	2.372
t = B/SE = "Significance" Statistic	Adjusted	2.371	4.405	1.877	3.388	0.125	-0.002	1.250	-0.392
	Unadj.	1.536	3.643	1.518	2.734	0.122	-0.377	0.630	-0.412
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.673	1.397	0.618	1.066	0.045	-0.0005	0.554	-0.140
	Unadj.	0.598	1.374	0.600	1.031	0.050	-0.161	0.274	-0.170
N = Number of Schools in Computation	FT	11	11	11	11	11	11	11	11
	NFT	10	10	10	10	10	10	10	10

Figure VII - 16
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 8

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:
Cov. Adj. Effect
Unadj. Effect

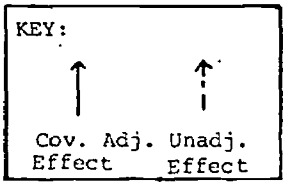


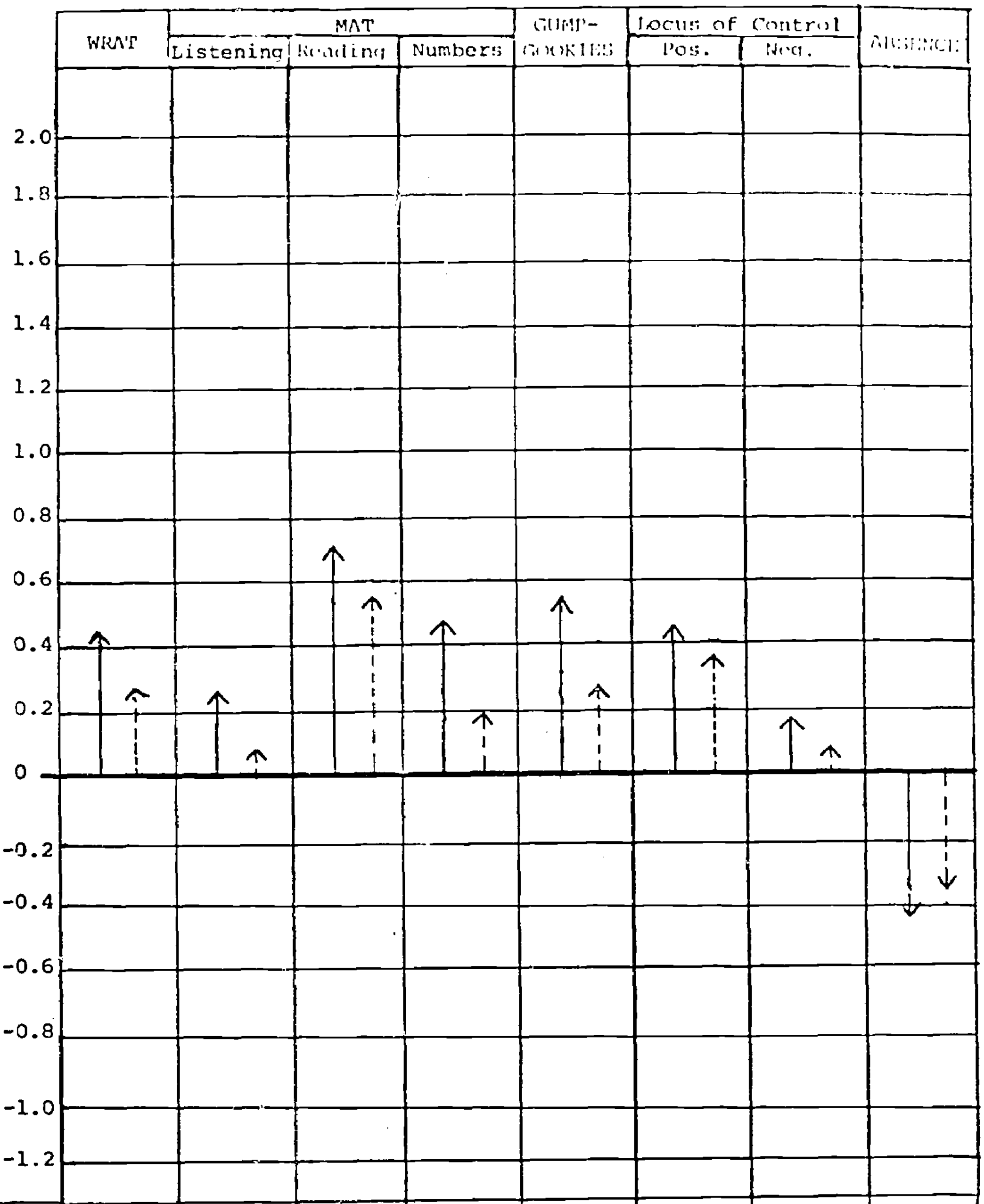
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.934	3.416	1.749	4.333	9.934	-0.281	-0.071	-0.210
	Unadj.	1.372	4.808	2.876	5.478	9.515	-0.143	-0.092	-2.085
SE = Standard Error of B	Adjusted	0.195	0.919	0.624	0.760	3.487	0.157	0.122	1.625
	Unadj.	0.269	1.097	0.752	0.915	3.980	0.190	0.120	1.855
t = B/SE = "Significance" Statistic	Adjusted	4.790	3.716	2.804	5.700	2.849	-1.790	-0.586	-0.129
	Unadj.	5.101	4.382	3.825	5.987	2.391	-0.752	-0.763	-1.124
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	1.057	0.918	0.718	1.396	0.793	-0.495	-0.202	-0.036
	Unadj.	1.552	1.292	1.181	1.765	0.759	-0.251	-0.261	-0.362
N = Number of Schools in Computation	FT	20	20	20	20	20	20	20	20
	NFT	15	15	15	15	15	15	15	15

Figure VII - 17
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 9

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



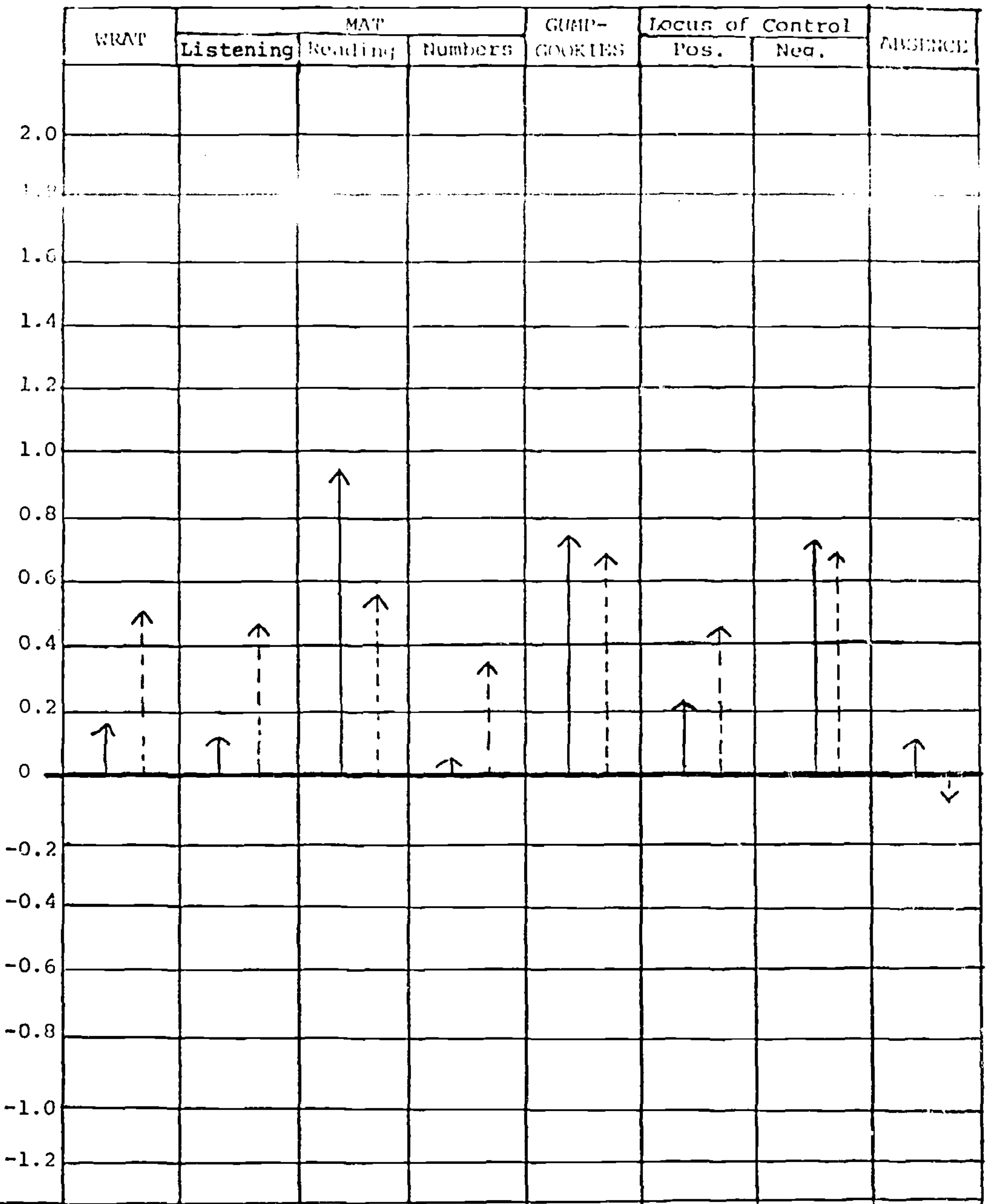


		WRAT	Listening	Reading	Numbers	GUMP-GOOKIES	Pos.	Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.387	0.993	1.718	1.436	7.107	0.252	0.066	-2.385
	Unadj.	0.244	0.313	1.401	0.610	3.164	0.214	0.024	-2.090
SE = Standard Error of B	Adjusted	0.239	1.123	0.762	0.929	4.261	0.192	0.149	1.985
	Unadj.	0.329	1.341	0.919	1.118	4.864	0.232	0.147	2.266
t = B/SE = "Significance" Statistic	Adjusted	1.619	0.884	2.254	1.546	1.668	1.312	0.444	-1.201
	Unadj.	0.743	0.233	1.525	0.545	0.651	0.922	0.162	-0.922
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.438	0.267	0.706	0.463	0.567	0.444	0.188	-0.414
	Unadj.	0.276	0.084	0.576	0.196	0.253	0.377	0.068	-0.363
N = Number of Schools in Computation	FT	11	11	11	11	11	11	11	11
	NFT	12	12	12	12	12	12	12	12

Figure VII - 18
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 10

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



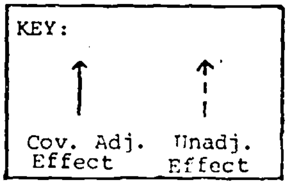
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.150	0.551	0.711	0.155	9.569	0.132	0.262	0.695
	Unadj.	0.463	1.743	1.421	1.139	8.498	0.253	0.252	-0.324
SE = Standard Error of B	Adjusted	0.237	1.112	0.755	0.920	4.220	0.190	0.147	1.966
	Unadj.	0.325	1.324	0.907	1.032	4.803	0.229	0.145	2.238
t = B/SE = "Significance" Statistic	Adjusted	0.636	0.495	0.942	0.168	2.268	0.693	1.782	0.353
	Unadj.	1.426	1.316	1.566	1.103	1.769	1.106	1.735	-0.145
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.170	0.148	0.942	0.050	0.763	0.232	0.745	0.121
	Unadj.	0.524	0.468	0.584	0.367	0.678	0.446	0.716	-0.056
N = Number of Schools in Computation	FT	17	17	17	17	17	17	17	17
	NFT	9	9	9	9	9	9	9	9

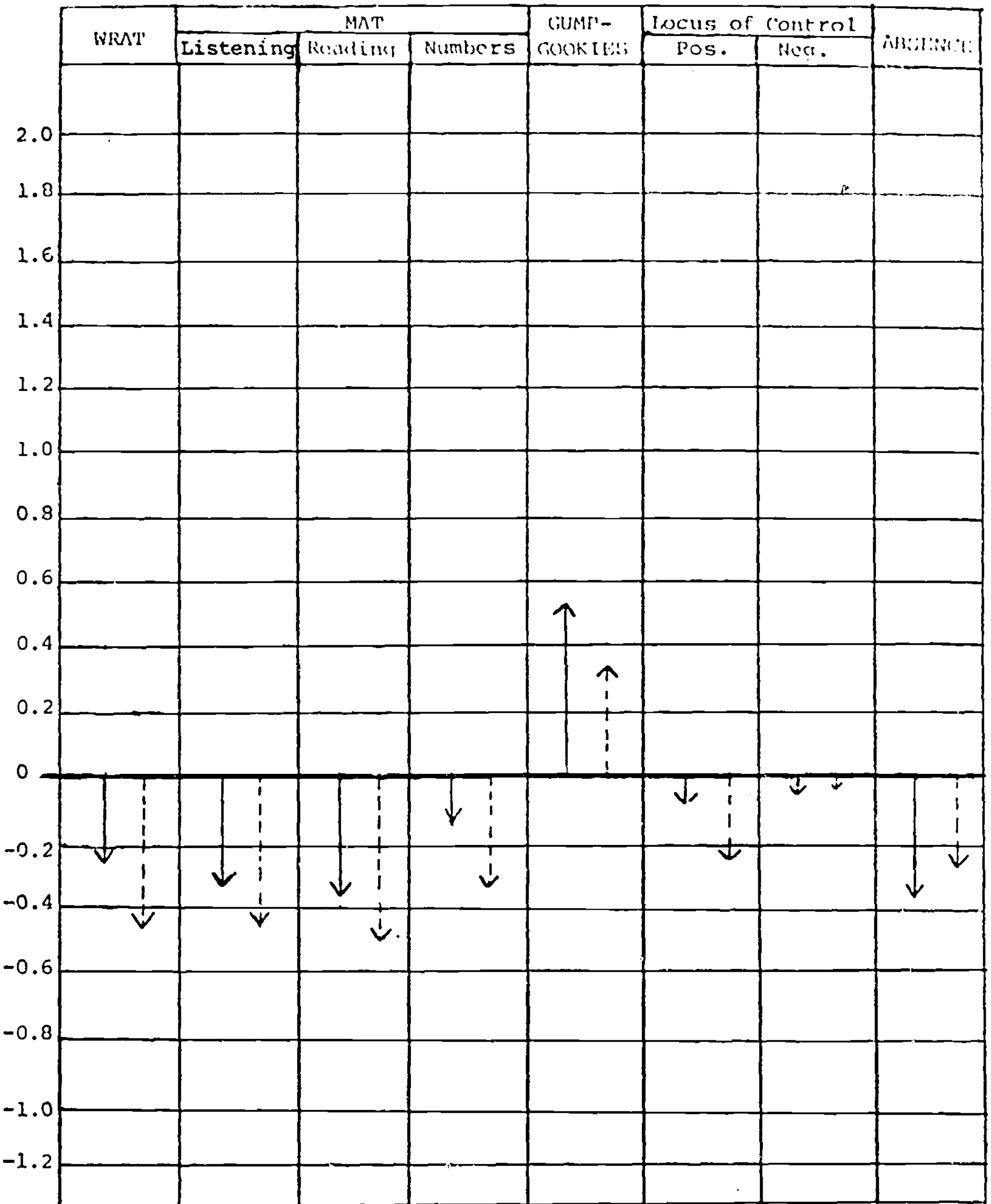
Figure VII - 19

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 11

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)





		WRAT	Listening	Reading	Numbers	GUMP-COOKIES	Pos.	Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.205	-1.274	-0.901	-0.435	6.775	-0.042	-0.014	-2.162
	Unadj.	-0.382	-1.645	-1.153	-0.993	4.469	-0.143	-0.009	-1.482
SE = Standard Error of B	Adjusted	0.225	1.057	0.717	0.874	4.008	0.181	0.140	1.868
	Unadj.	0.315	1.286	0.881	1.072	4.665	0.222	0.141	2.174
t = B/SE = "Significance" Statistic	Adjusted	-0.914	-1.206	-1.257	-0.498	1.690	-0.230	-0.099	-1.158
	Unadj.	-1.211	-1.279	-1.308	-0.926	0.958	-0.643	-0.063	-0.682
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	-0.232	-0.342	-0.370	-0.140	0.541	-0.073	-0.039	-0.375
	Unadj.	-0.432	-0.442	-0.474	-0.320	0.357	-0.252	-0.025	-0.257
N = Number of Schools in Computation	FT	13	13	13	13	13	13	13	13
	NFT	12	12	12	12	12	12	12	12

Figure VII - 20

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 14

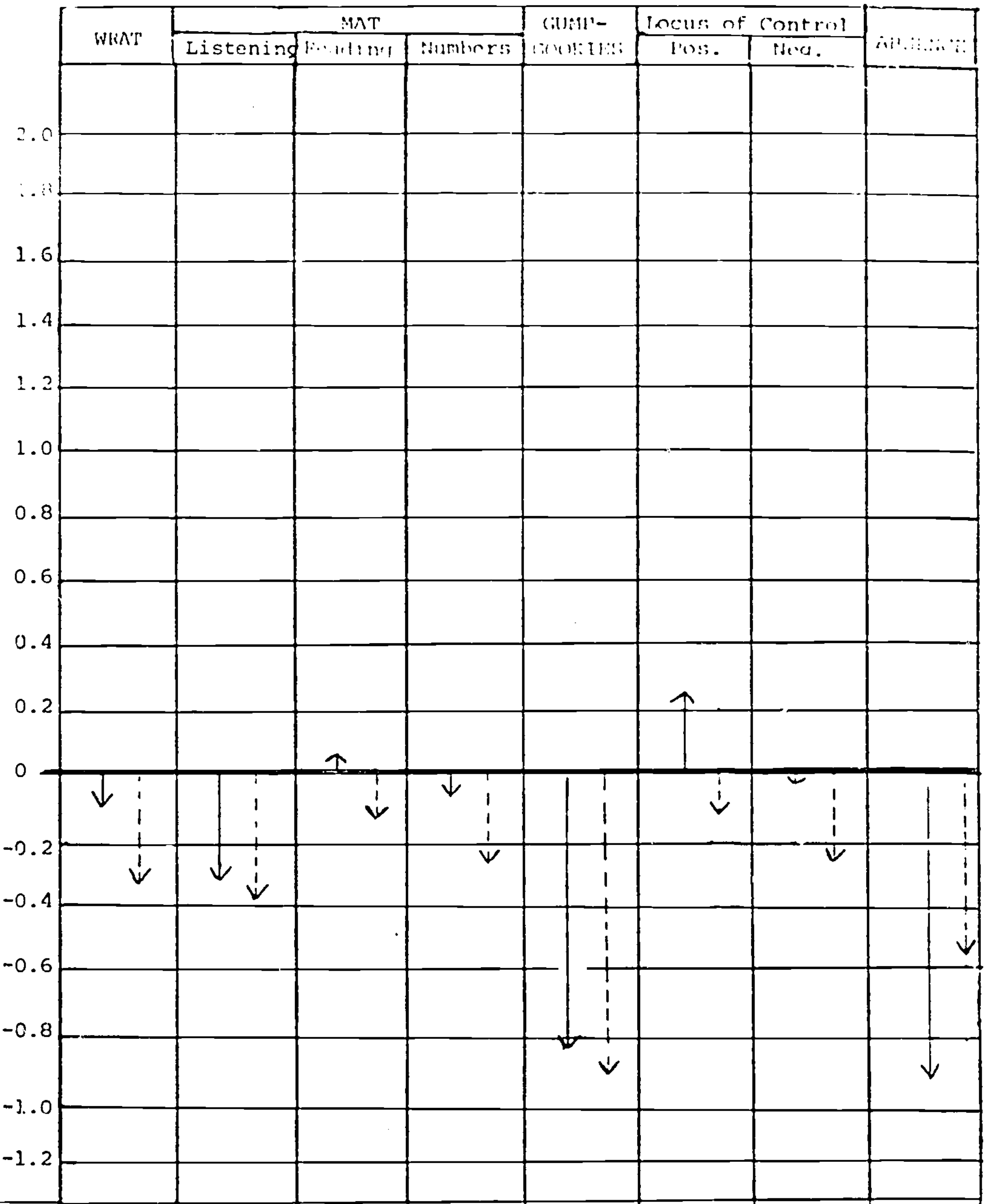
Including [illegible]

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

↑ Cov. Adj. Effect

⇡ Unadj. Effect



		WRAT	Listening	Reading	Numbers	GUMP-GOOKIES	Pos.	Neg.	
B = Magnitude of Effect	Adjusted	-0.081	-1.191	0.145	-0.159	-10.222	0.142	-0.007	-5.[illegible]01
	Unadj.	-0.294	-1.452	-0.276	-0.731	-10.955	-0.052	-0.081	-3.310
SE = Standard Error of B	Adjusted	0.270	1.271	0.863	1.051	4.823	0.218	0.168	2.24[illegible]
	Unadj.	0.383	1.561	1.070	1.301	5.662	0.270	0.171	2.639
t = B/SE = "Significance" Statistic	Adjusted	-0.300	-0.937	0.168	-0.151	- 2.119	0.653	-0.044	-2.314
	Unadj.	-0.768	-0.930	-0.259	-0.562	- 1.935	-0.192	-0.472	-1.255
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S - Effect in Standard Deviations	Adjusted	-0.091	-0.320	0.059	-0.051	- 0.816	0.250	-0.020	-[illegible].903
	Unadj.	-0.333	-0.390	-0.114	-0.236	- 0.874	-0.091	-0.229	-0.575
N = Number of Schools in Computation	FT	9	9	9	9	9	9	9	9
	NFT	8	8	8	8	8	8	8	8

3.0 RELATED DATA

In this section, we will examine two sets of data which add to our understanding of the FT/NFT contrasts. The first is a set of correlations relating pre- and posttest scores with measures of socioeconomic status (SES). The second [illegible] study which was introduced earlier in this [illegible].

3.1.0 SOCIOECONOMIC STATUS AND PUPIL ACHIEVEMENT

Another approach to the overall effects of the FT programs is to consider the relationships between indicators of social status and achievement for the FT group at the beginning of the kindergarten year and then again at the end of the year. If these FT programs are having an effect, it would be expected that the contribution of social status to achievement would diminish. Equalization of educational opportunities means that the distribution of achievement is not influenced by the social status of children or their families.

In order to examine this issue, the zero order correlations of three social status indicators with the Fall and Spring WRAT are presented in Table VII-8 for both FT and NFT groups at the school level of analysis.

First, it is clear that the correlations between SES and Fall WRAT are lower for the FT group than for the NFT group of schools. Analyses have not yet been run to determine the reasons for these differences.[1] The hypothesis that will guide the search for the basis of the differences is that children in the FT group already show some treatment effects from their preschool experience, which is rather more extensive than the preschool experiences acquired by the NFT group. The comparison between NFT and FT rates of preschool is directly available at the child level of aggregation and these data indicate that 72% of the FT

---

[1]Table III-3 indicates that the variances of SES indicators, Fall WRAT and Spring WRAT, are essentially the same in FT and NFT. Attenuation via restricted variability could account for the differences in these covariates.

Table VII - 8

ZERO ORDER CORRELATIONS BETWEEN SES INDICATORS AND FALL AND SPRING WRAT

FT Schools
N = 156

	Fall WRAT		Spring WRAT		P Differences
	r	r^2	r	r^2	
Adjusted income level	.4675	.2185	.2931	.0859	<.025
% Minority	-.2985	.0891	-.1132*	.0128	<.025
Mother's education	.4748	.2254	.3309	.1095	<.025

NFT Schools
N = 132

	r	r^2	r	r^2	
Adjusted income level	.7615	.5799	.6034	.3641	<.001
% Minority	-.6216	.3864	-.5195	.2698	<.025
Mother's education	.7213	.5202	.5305	.2811	<.001

* This correlation is significant at the .08 level. All other correlations are significant at the .01 level.

children had preschool experience prior to kindergarten while only 45% of the NFT had such experiences. Further examination of these data is necessary to test this hypothesis.

Next, the reduction in correlations from the Fall to the Spring WRAT for the FT group is also apparent. Tests of the significance of the differences between correlations of SES measures and Fall and Spring WRAT scores utilizing Fisher's z transformation indicate that all Spring WRAT correlations are significantly lower than Fall WRAT correlations at the 0.02 level. It is clear that something has happened in the kindergarten year to influence these correlations. However, it is also true that the correlations between SES and the Fall and Spring WRAT scores for the NFT are also significantly reduced in the same direction. It should be noted, on the other hand,that the comparison of the reductions in these differences is not a reasonable test of the hypothesis. The NFT correlations are so much higher than the FT correlations that the differences in the standard errors of the sets of correlations make it considerably easier for small differences in NFT correlations to reach significant levels. Although it is clear that something has happened to the NFT group over the kindergarten year to reduce the correlations (and this may have to do with the special federally funded programs such as Title I which are so often present in these schools), what is impressive are the differences in the variance of the posttest scores accounted for by the SES factors in the FT and NFT groups. Whereas in the NFT group at the end of kindergarten, 36% of the WRAT score variance is accounted for by adjusted family income, only 9% of the WRAT variance is accounted for by adjusted family income in the FT group. Similarly, 27% of the variance of posttest scores are accounted for by the percent of minority children in the school for the NFT groups, whereas this figure is 1% for the FT group. Finally, mother's education accounts for 28% of the Spring WRAT variance in the NFT group, and only 11% of the WRAT variance is accounted for by this variable in the FT group. Because of the lower SES/Fall WRAT correlations in the FT group, it is reasonable to conclude that these relatively smaller portions of the posttest variance accounted for by the SES

variables is a result of the accumulation of the greater amounts of preschool experience and the differential kindergarten experiences of the FT group.

Clearly, zero order correlations cannot be taken as definitive evidence for the reduction of SES influence on achievement in the FT groups. However, these data can be taken as supporting evidence that FT, in the aggregate, is providing positive effects.

3.2.0 TIME OF TESTING CORRELATIONS

Additional information on each Sponsor's FT effects is found in a study of how the Spring test scores vary over the posttest interval. Are Spring test scores related to the length of the instructional interval? One answer to this question was established by correlating the number of days between pre- and posttest (the length of the instructional interval) with the posttest scores, partialling out Fall WRAT scores. By controlling for initial WRAT scores we hope to remove some of the problems introduced by a non-random testing schedule; in some instances schools with higher Fall WRAT scores were being tested later in the Spring thereby unjustly increasing the zero-order correlation between length of instructional interval and posttest scores. The reader is reminded that the partial correlations reported only reflect the extent to which the scores are related to the time interval covered by posttest administration. That is to say, with the present data we have no measure of what effect Sponsors are having on the outcome variables in the time period from the last administration of the pretest to the first administration of the posttest for any given Sponsor. Care should be taken in not extending the relationships presented below to the range of the instructional interval not included in the values of the time of testing variable.

Tables VII-9 and VII-10 present the partial correlations of length of instructional interval with Spring scores for each Sponsor's FT and NFT groups respectively. Correlations significant at the .10 level (using a two-tailed test) are starred. Those correlations which are not significantly different from zero imply that there is no simple relationship between the length of the instructional interval and the Spring scores

TABLE VII - 9

Partial Correlations of Spring Scores
with Length of Instructional Interval
Controlling for Fall WRAT
FT Schools [a]

SPRING SCORES	SPONSOR 2 N=29	3 N=21	5 N=16	7 N=11	8 N=19	9 N= 9	10 N=16	11 N=13	12 N= 9	14 N= 9
Spring WRAT	.1323	.3163	.0088	-.2209	-.0204	.2582	-.0599	.6423*	.3079	.0579
MAT Sounds	.0544	.6057*	.5692*	-.0032	.3148	.2852	-.1214	.6221*	.2285	-.5663
MAT Read	.2283	.4544*	.2010	-.5206	.0972	.0925	.2267	.5221*	.6158	-.2807
MAT Arith.	.3068	.5833*	.7101*	-.0934	.0649	-.0301	.4671*	.6600*	.2174	.1143
GUMP-GOOKIES	.1497	-.7108*	.5073*	.3071	.6052*	.3788	.6040*	.4411	.0856	.7793*
LOCUS OF CONTROL Positive	.3432*	.1567	-.0031	.1377	-.0475	.4150	-.5531*	.4817	-.6971*	-.5531
LOCUS OF CONTROL Negative	.0759	.3306	.3124	.3586	.1839	.6656*	.0088	.1283	-.6746*	.0856

*These correlations are significant at the .10 level, using two-tailed significance tests.

[a]The sample of FT schools for this study was 152, four schools less than the total analytic subset. These four schools are not included here because of insufficient time of testing information.

Table VII - 10

Partial Correlations of Spring Scores
with Length of Instructional Interval
Controlling for Fall WRAT
NFT Schools[a]

SPRING SCORES	SPONSOR 2 N=20	3 N=20	5 N=14	7 N=10	8 N=15	9 N=10	10 N= 9	11 N=12	12 N=11	14 N= 8
Spring WRAT	.1153	.3198	.2631	.0953	.3128	-.4061	.4329	.1477	-.6257*	-.3440
MAT Sounds	-.2257	.0496	.4394	.5301	.2649	-.1196	.1029	.4838	-.4007	.0502
MAT Read	-.0928	.1080	-.0364	.2351	.1904	-.0562	.0225	.0378	-.7641*	.3610
MAT Arith.	.1429	-.1583	.6774*	.2500	.3982	.1862	.3014	.6282*	.4723	.0114
GUMP-GOOKIES	.1825	-.1542	.4578	-.4549	.6929*	-.5783	.8709*	.6790*	-.3429	-.4971
LOCUS OF CONTROL Positive	-.0588	.2706	-.1933	-.1067	-.3239	.1280	-.4619	-.0835	.6388*	.2658
LOCUS OF CONTROL Negative	.1312	.2773	-.0187	.1581	-.3930	-.1214	-.5339	-.1461	.5446	-.4558

*These correlations are significant at the .10 level, using two-tailed significance tests.

[a] The sample of NFT schools for this study was 129, 3 schools less than the total analytic subset. These three schools are not included here because of insufficient time of testing information.

after adjusting for pretest scores; positive correlations indicate that after adjusting for pretest scores, the longer the instructional interval the higher the posttest score; negative correlations indicate that after adjusting for pretest scores, the longer the instructional interval the lower the posttest score.

Sponsor 2 (Far West) has only one outcome measure, Locus of Control (positive), which correlates positively with the length of the instructional interval.

For Sponsor 3 (University of Arizona) there are positive correlations between the length of the instructional interval and all MAT subtest scores, but a negative correlation for the Gumpgookies test.

Sponsor 5 (Bank Street) has positive correlations between the time of testing variable and MAT Listening to sounds, MAT Arithmetic, and Gumpgookies. The NFT comparison schools for Bank Street also has a positive correlation between instructional time and Arithmetic.

After adjusting for pretest scores, University of Oregon (Sponsor 7) has no significant correlations between the outcome measures and length of the instructional interval.

For University of Kansas (Sponsor 8) and SEDL (Sponsor 14) there are positive correlations between instructional time and achievement motivation but no significant correlation between achievement and instructional time. The NFT schools for Kansas also have a positive correlation here but the comparison schools for SEDL do not.

An achievement measure (MAT Arithmetic), the achievement motivation measure, and a Locus of Control measure (positive) all correlate significantly with the length of the instructional interval for University of Florida (Sponsor 10). Locus of Control (positive) has a negative correlation, while Arithmetic and Gumpgookies have positive correlations. Only the positive correlation between the Gumpgookies score and the time of testing variable is found in the NFT group.

EDC (Sponsor 11) has positive correlations between all four of the achievement measures and length of the instructional interval. The NFT comparison schools for EDC have positive correlations for MAT Arithmetic and Gumpgookies.

University of Pittsburgh (Sponsor 12) has negative correlations with both Locus of Control measures. The NFT comparison schools have a positive correlation for Locus of Control (positive), and negative correlations for Spring WRAT and MAT Reading.

The conclusion drawn from these time of testing correlations is consistent with some other results seen thus far: namely, Sponsors produce varied results. Some Sponsors seem to have a positive effect on the scores of the Spring tests across the interval represented by instructional time; other Sponsors have no effect; still others have a negative effect on these scores. Many questions come to mind based on the results presented above. Why don't the structured programs, such as University of Oregon, U. of Kansas, and U. of Pittsburgh, have positive correlations on the achievement measures while the open classroom approaches of Bank Street and EDC do? Why does the University of Arizona have positive correlations on the achievement measures and a negative correlation on achievement motivation? Can we explain the correlations produced by the NFT schools?

The small N on which these correlations are based, the unavailability of test scores across the full school year, and insufficient information on programs operating within the NFT schools force us to leave these questions open for future studies. Further, it would be wise to attempt replication of such correlational data before attempting to account for them. For the present it is clear that no simple relationship exists between the several testing intervals and the scores generated during those intervals. Sponsors who expect such relationships because of the nature of their models must be considered in the light of more data than are available at this writing in order to fully test their expectations.

4.0 SPONSOR VIGNETTES

4.0.1 INTRODUCTION

In order to provide a picture of the contribution of each Sponsor to the overall FT/NFT contrasts presented above, the following section presents the effects produced by the Sponsors individually. A short narrative putting these effects in the context of the particular children, sites, and other demographic data involved in these school level contrasts is also provided.

To add further to the understanding of the effects for each Sponsor, the individual main effects of class and child level studies will also be presented for each Sponsor.[5] These studies were designed to answer specific questions about the effects produced by each Sponsor when working with types of classes and types of children. The results of these interaction studies are presented in Chapter VIII, but in the course of examining interactions, individual Sponsor main effects are also produced. These are presented here as a way of providing multiple approaches to the question of Sponsor effects.

The sites selected for participation in the national evaluation were not designed to be representative of the sites with which each Sponsor has been working. Thus, it is important that each Sponsor know the particular schools included in these analyses in order to assess the representativeness of the findings. The specific schools are not presented here for reasons of space and the protection of school anonymity, but they are summarized by geographic region in sufficient detail so that each Sponsor should be able to recognize which of the full set of participating schools are present in this summary. All schools selected by USOE for inclusion in the national evaluation, and for which a full set of data were available, were included in these analyses.

Class and child level studies were based upon a subset of classes and children included in the analytic group of schools. A somewhat different set of inclusion criteria were applied to classes and children for those studies, so that the subsets produced at class and child levels are different from the school level groups both in numbers and characteristics. The divergence of these multiple approaches are discussed in these summaries of each Sponsor's effects.

---

[5] The methods used in the class and child levels of analyses are presented in Sections VII: 1.2 and VII: 3,2 respectively.

It must be remembered that these vignettes are not to be taken as full descriptions of Sponsor effects. The focus is on the impacts the Sponsors' have had on a diverse group of children scattered over a variety of sites, under a wide variety of conditions, all of which are combined in a single analysis for each Sponsor. In addition, only a selection of contextual factors have been included in the narrative, simply to provide a sense of the range of conditions contributing to the single set of numbers for each Sponsor.

Before presenting these vignettes, the bases of interpretation of the three levels of findings (school, class, and child) must be considered.

The present section deals with the effects of Sponsors' programs at the school, class, and child level; that is, the difference between FT and NFT groups within Sponsors at each level is presented. Along with a presentation of the results an attempt is made to highlight the characteristics of a Sponsor's approach and sample that make both his program and his schools, classes, and children unique.

For each Sponsor, not all of the sites in which his program is operating are represented in the study. The geographic data should allow Sponsors to judge whether at any given level they are fairly represented. Demographic and background characteristics are also presented whenever an outstanding characteristic appears. The details of geographic, demographic, and background characteristics are presented in Chapter III.

For each level of analysis, the results are presented as contrasts between the FT and NFT groups. Positive valued contrasts ("+" sign with arrow facing upward) represent FT-favoring results and negative valued contrasts ("-" sign), represent NFT-favoring results, except on the Absence outcome where the opposite is true. It is important to keep in mind the intervention of the Sponsor when interpreting these results. Some approaches do not attempt to produce achievement gain in early years, while other programs do. Each vignette begins with a brief description of the Sponsor's intent so that the reader can keep it in mind in examining the outcome patterns.

At the school and class levels of analysis, an adjusted and an unadjusted difference between the FT and NFT groups are presented, and at the child level an unadjusted difference and a pair of adjusted differences are provided. The unadjusted difference represents the raw difference between the FT and NFT scores on the outcome. The adjusted difference represents the difference between FT and NFT scores when initial inequalities, differences on initial achievement level, SES factors, and geographic location are compensated for. In the child level profile a true score adjusted contrast is also presented. This contrast represents the difference between FT and NFT when the difference has been adjusted for initial inequalities, as well as the unreliability of the most important and fallible covariable, initial achievement level as measured by the Fall WRAT.

In general the statistical significance of the results is ignored except at the school level, where the smallest number of units are utilized in the analysis and t tests are presented. For the class and particularly the child level, the statistical significance of the contrasts adds little to our understanding of the results, since practically all contrasts are significant due to the large number of observations. For all levels of analysis, all results that are larger than a quarter of a standard deviation are presented for discussion. This criterion is admittedly statistically arbitrary; however, a quarter of a standard deviation is perhaps an appropriate index of educational impact and does provide an heuristic for decision making.

Before considering the Sponsors, let us explore the levels of analysis and reiterate their purpose.

4.0.2 LEVELS OF ANALYSIS

4.0.2.1 School Level

The school is an important unit of analysis both systemically and statistically. Both FT and NFT schools were selected because they displayed certain SES characteristics. The decision as to what Sponsor's program would be applied in a school often involved school personnel and parents. The Follow Through program is in part administered at the school level, that is, the nutritional, medical and support services

associated with Follow Through are applied or available at the school level. Furthermore, there are other aspects of the Follow Through program that occur on a school wide basis, e.g., the Policy Advisory Committee and, in some cases, Sponsor training.

On statistical grounds, the school represents a more stable unit than the smaller units and is more free of measurement error and idiosyncrasies of particular children, parents, teachers, and classes. Furthermore, since the school, in an important sense, represents the experimental unit to which the "treatment" is applied, smaller units of analysis within the school lack independence both in a practical and statistical sense. Teachers, parents, and pupils within the same school interact and the interaction is an essential part of the treatment. As such, class and child level analyses are likely to amplify the school level effects in a biased manner depending on the ratio of classes per school, and children per school (Porter, 1972).

While the school is a legitimate unit of analysis, there are certain limitations inherent in aggregate measures concerning the nature of inference that they permit. Effects at the school level say little about benefits or deficits accrued by particular types of children or classes. An effect at the school level could result from any of a variety of confoundings within the school. For example, higher SES children in a school may benefit substantially from a program and leave the lower SES children far behind. The aggregate of the scores of the children, the school mean, could indicate a gain that is biased in such a manner. The likelihood of uniform confounding across many of a Sponsor's schools is low; however, the efficacy of a program cannot be based on a single level of analysis.

The school level study is thus addressed to the questions:

- What are the particular school effects of particular Sponsors?
- What is the variability of these effects?
- With what kind of geographic distribution?
- With what kind of initial differences between the Follow Through and non-Follow Through groups?

4.0.2.2 Class Level

Many of the arguments that apply to the variability of the school level of analysis apply to the class level as well. The class represents a level of application of the Follow Through approach, that is, the children, their parents, the teachers, the materials, and the instructional program that represents a Follow Through approach all come together and interact in the classroom. The classroom is a model-relevant unit of analysis in that the constituents of a Sponsor's approach often have their most intimate contact and interaction at this level.

On statistical grounds, the class aggregate values have the virtue of stability and some measure of independence. Class aggregates are relatively free of measurement error, although they are likely to amplify tester related error, at least to the extent that tester error is biased and unevenly distributed across tested classes. Classes are also independent inasmuch as the Follow Through approach can be viewed as a classroom treatment.

As in school level analyses, inferences from class to child level results are limited since the treatment may interact with child characteristics. Similarly, inference to the school level is obviated by the possibility of the interaction of treatment and class characteristics such as initial ability level or ethnic composition. Furthermore, there may be biases in the way in which classes are aggregated to schools. For example, if most classes showing gains are in a small number of schools, a class effect may not be reflected in the school aggregate.

The class level study is addressed to the questions:

- What are the particular class level effects of particular Sponsors?
- With what kind of geographic distribution?
- With what kind of initial differences between Follow Through and non-Follow Through groups ?

4.0.2.3 Child Level

The child level is perhaps the most difficult to justify on statistical grounds. Problems of measurement error abound. Scores at this level may amplify effects of testing conditions, tester biases,

pupil selection bias related to who gets tested when as well as who gets into the Follow Through program in a school, etc. However, the differential effects of different programs on different kinds of children, must be addressed with the child as a unit of analysis. In addition, the general benefit of the Follow Through program on children can only be addressed here. Higher levels of analysis do not answer the question of whether a Follow Through program is maintaining the status quo of public education and giving benefits to children with certain restricted background characteristics or whether Follow Through is truly innovative and benefits those for whom compensatory education is intended.

The child level study thus addresses the questions:

- What are the particular child level effects of particular Sponsors?
- With what kind of geographic distribution?
- With what kind of initial differences between Follow Through and non-Follow Through?

Let us now turn to the results at each level of analysis for each Sponsor.

4.0.3 METHODS

The analytic designs for the school level of analyses were presented in full in section 1.2.3 of this chapter. At the class and child level of analysis the treatment within Sponsor effects reported were calculated using the nested design.

4.1.0 SPONSOR 2: FAR WEST LABORATORY, RESPONSIVE EDUCATIONAL PROGRAM

The Sponsor describes the Responsive Educational Program as autotelic; that is, it is based on the philosophy that the best way for a child to learn is for him to explore and make discoveries from the environment around him. The responsive classroom environment is designed to help the child develop problem solving abilities, develop confidence in his own capacity to succeed, and develop the academic skills necessary for effective problem solving. While no single learning theory or method is applied, the model offers a variety of games, materials, and learning tasks to aid in the development of reasoning abilities and self-directed, self-rewarding behavior.

4.1.1 School Level FT/NFT Contrasts

The subset of schools meeting the criteria for inclusion in the school level analysis was drawn from one half of Far West's sites. Approximately 48% of the FT schools in this subset and 40% of the NFT schools are located in two medium-sized Western cities.

First we shall compare the subset of Far West's FT schools to the total group of FT schools analyzed for all Sponsors. The FT schools for this Sponsor are similar in respect to socioeconomic status to the average FT school for all Sponsors. That is, the mean adjusted income and the mean percentage of mothers completing high school for these schools are not markedly different from those of the FT schools for all Sponsors in the analytic sample. The FT schools are also close to the overall mean of entering achievement level for all Sponsors, as measured by the Fall WRAT.

Next, the FT and NFT schools participating in the Far West Laboratory program are contrasted. The mean adjusted income for the families of children attending the FT schools associated with the Far West Laboratory is somewhat lower than the mean adjusted income for the families of children in this Sponsor's NFT schools. However, the mothers of the children in the FT schools have achieved, on the average, the same educational level as the NFT mothers. At the same time, the mean

percentage of non-White pupils in the FT schools (68%) is higher than in the NFT schools (59%). Finally, the children in the FT schools entered kindergarten at essentially the same achievement level as the children in the NFT schools, so that at this level of analysis, we can conclude that a relatively close match between the FT and NFT groups has been achieved.

Figure VII-2 presents the profile of FT/NFT contrasts on Spring measures for the Far West Lab program at the school level of analysis. When the initial differences between the groups are partialled out there is only one significant FT/NFT contrast: the FT group exceeds the NFT group on the Gumpgookies test. In addition, there is a trend in favor of the FT group on the MAT reading subtest, and the variability of the MAT-Arithmetic subtest results across schools suggests that some FT schools may also be having positive effects in this area.

4.1.2 <u>Class Level FT/NFT Contrasts</u>

The group of classes which were included in these analyses were selected from two Western sites and three non-Western sites, which parallels the distribution of schools.

The mean adjusted income level for FT classes is somewhat lower than the mean for NFT classes which parallels the school district. bution. Mothers' educational levels are the same for the FT and NFT classes, and there is almost the same distribution of non-White children at class level as reported for school level (FT = 70%; NFT = 50%). Finally, the FT and NFT children have essentially the same mean entering achievement scores. On the whole, the differences between the FT/NFT children participating in the class level analysis appear to be similar to those found at the school level.

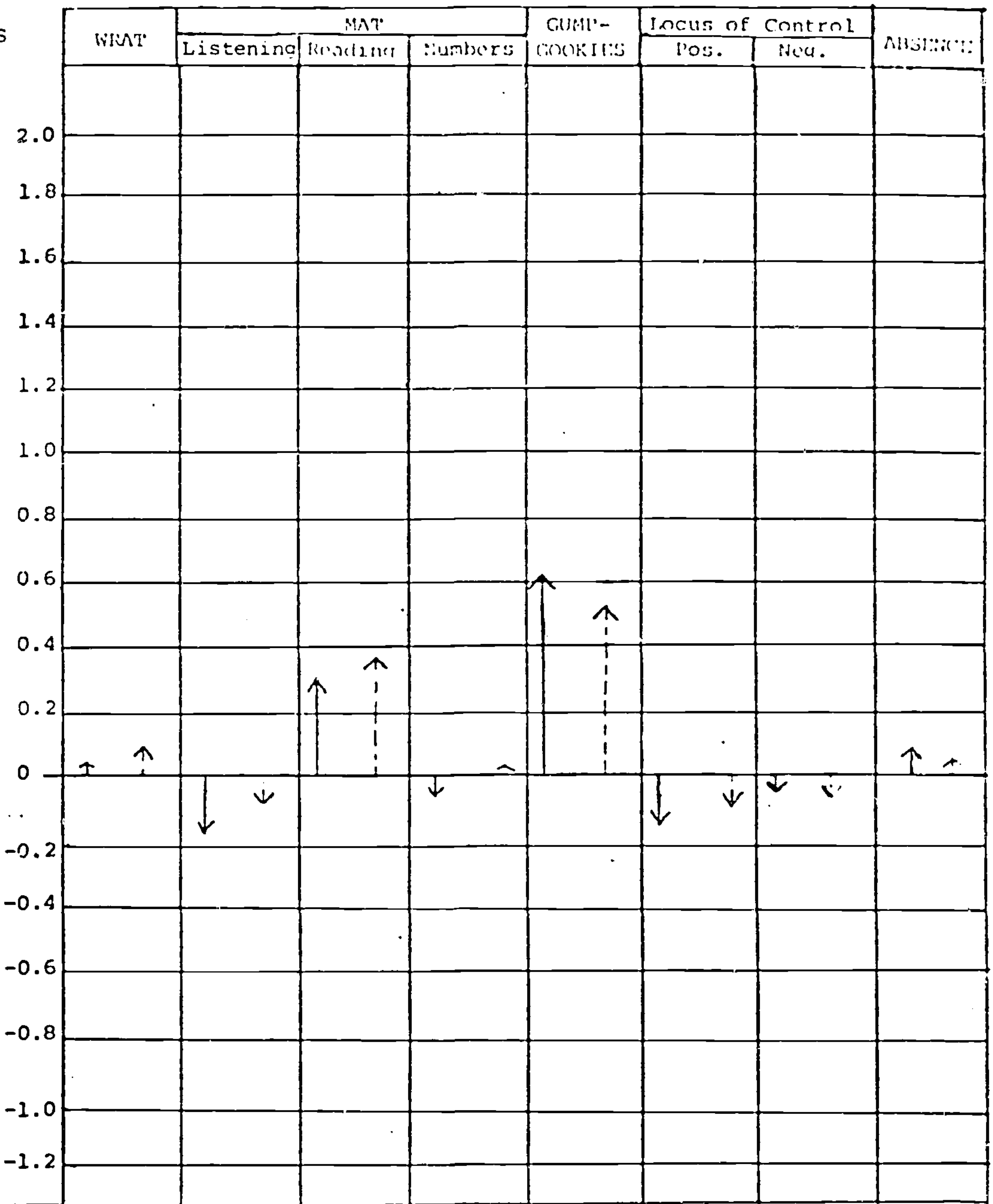
At this level, two important contrasts emerge on Spring outcome measures. (See Fig. VII- 21) The FT classes exceed the NFT classes on MAT reading and MAT listening for sounds. The variability in MAT arithmatic found at the school level is still present at class level. However, there is no FT favoring contrast on the Gumpgookies test at this level as there was at the school level. Given the similarity between the groups of children at these two levels, and the similarity on achievement outcomes, this finding is not explicable with the current set of data.

Figure VII - 2

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 2

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



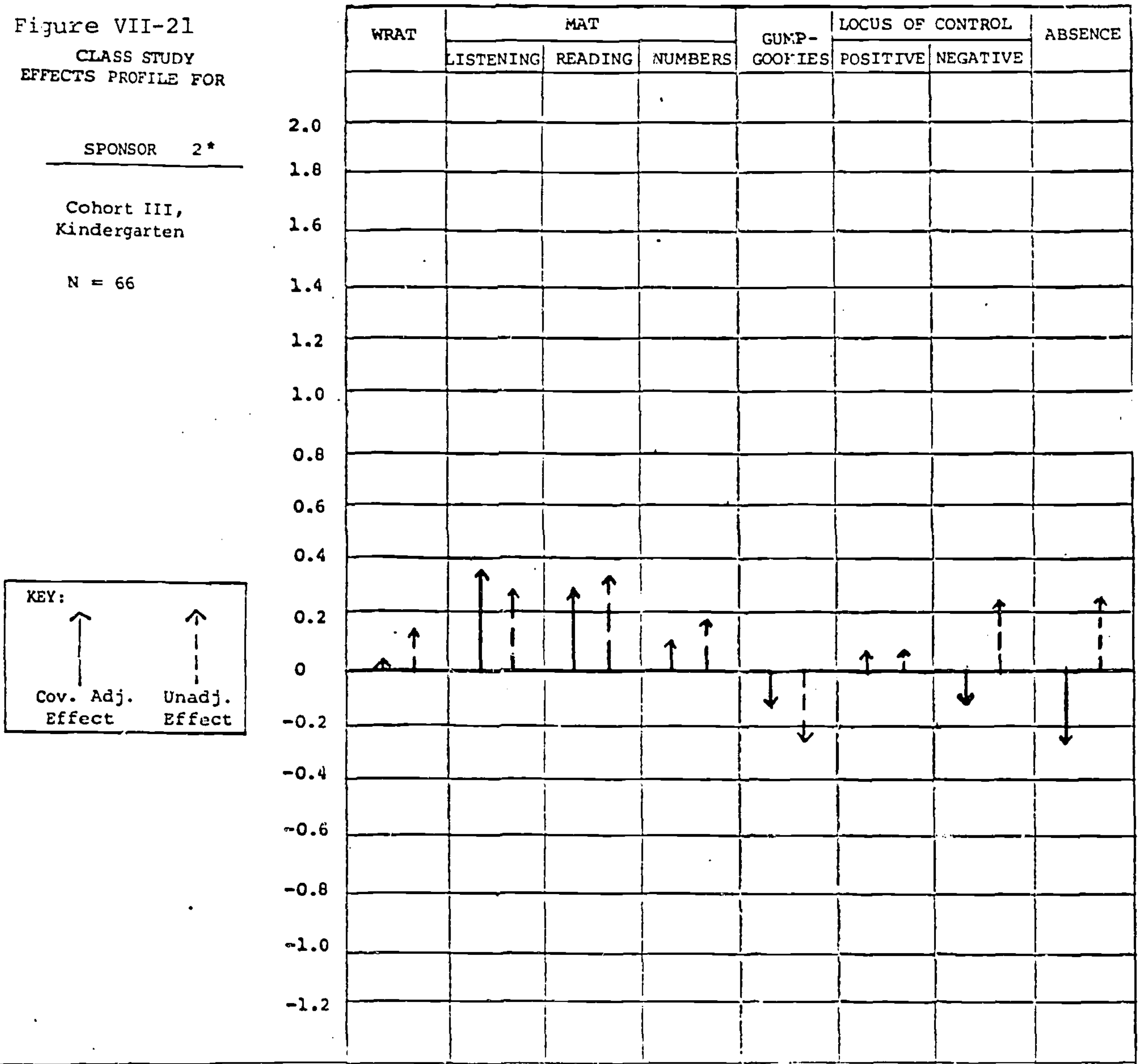
		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.020	-0.683	0.690	-0.201	6.847	-0.088	-0.016	0.415
	Unadj.	0.085	-0.301	0.938	0.039	6.086	-0.041	-0.022	0.219
SE = Standard Error of B	Adjusted	0.158	0.743	0.516	0.650	2.444	0.126	0.096	1.138
	Unadj.	0.229	0.936	0.649	0.804	2.743	0.157	0.096	1.363
t = B/SE = "Significance" Statistic	Adjusted	0.128	-0.919	1.338	-0.309	2.802	-0.698	-0.163	0.364
	Unadj.	0.371	-0.321	1.444	0.049	2.218	-0.261	-0.229	0.160
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.023	-0.182	0.283	-0.064	0.607	-0.160	-0.047	0.083
	Unadj.	0.097	-0.080	0.385	0.012	0.540	-0.074	-0.067	0.044
N = Number of Schools in Computation	FT	29	29	29	29	29	29	29	29
	NFT	20	20	20	20	20	20	20	20

Figure VII-21
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 2*

Cohort III,
Kindergarten

N = 66



		WRAT	LISTENING	READING	NUMBERS	GUMP-GOOKIES	POSITIVE	NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	.060	1.609	.831	.370	-.335	.070	-.061	-1.261
	Unadj.	1.093	1.382	1.105	.671	-.842	.046	-.121	-1.468
* B/S = Effect in Standard Deviations	Adjusted	.008	.368	.274	.099	-.098	.087	-.111	-.208
	Unadj.	.153	.316	.365	.180	-.248	.058	-.219	-.243
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

4.1.3 Child Level FT/NFT Contrasts

The demographic characteristics of the children participating in these analyses are essentially the same as those involved in the analyses reported thus far. Income levels are lower for the FT children, mother's education and entering achievement levels are the same for the two groups. Percent non-White for the FT group of children is 63% and for NFT children it is 43%, percentages which are very close to the figures for the children involved in the previous analyses.

At this level, there are no contrasts which indicate an advantage for the FT children, although the same trends are present in these data as found in the previous analyses. (See Fig. VII-22.) All of the achievement results are in the FT direction but none are impressive enough to discuss.

4.1.4 Selected Teacher Data

As described in the Teacher Monograph, Sponsor 2 is known to have the most experienced FT teachers in the sample. Not only have they taught longer overall than any other FT group, but they have been in their Sponsor's program longer. On the average they have been with the Far West Lab program for almost three years.

The kindergarten teachers report receiving relatively little training from their Sponsor, perhaps because they have recieved a great deal in previous years. However, their values reflect the philosophical orientation of the Responsive Educational Program. They value working with parents more than their NFT counterparts, despite the fact that the NFT group for this Sponsor is more highly parent-oriented than any other NFT group. They also make a great many home visits both relative to other FT teachers and compared to their NFT group. The FT teachers are as child-centered in goals and practices as the NFT group, and this NFT group is one of the most child-centered in the sample.

4.1.5 Summary and Discussion

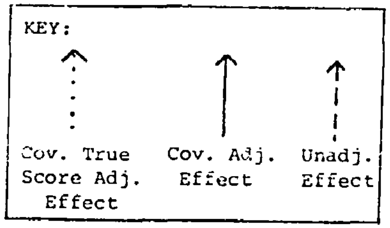
Far West Lab program appears to be having some impact on both achievement and motivation outcomes. At the school level of analysis, the program was found to have a significant positive effect on the development of achievement motivation, as measured by the Gumpgookies

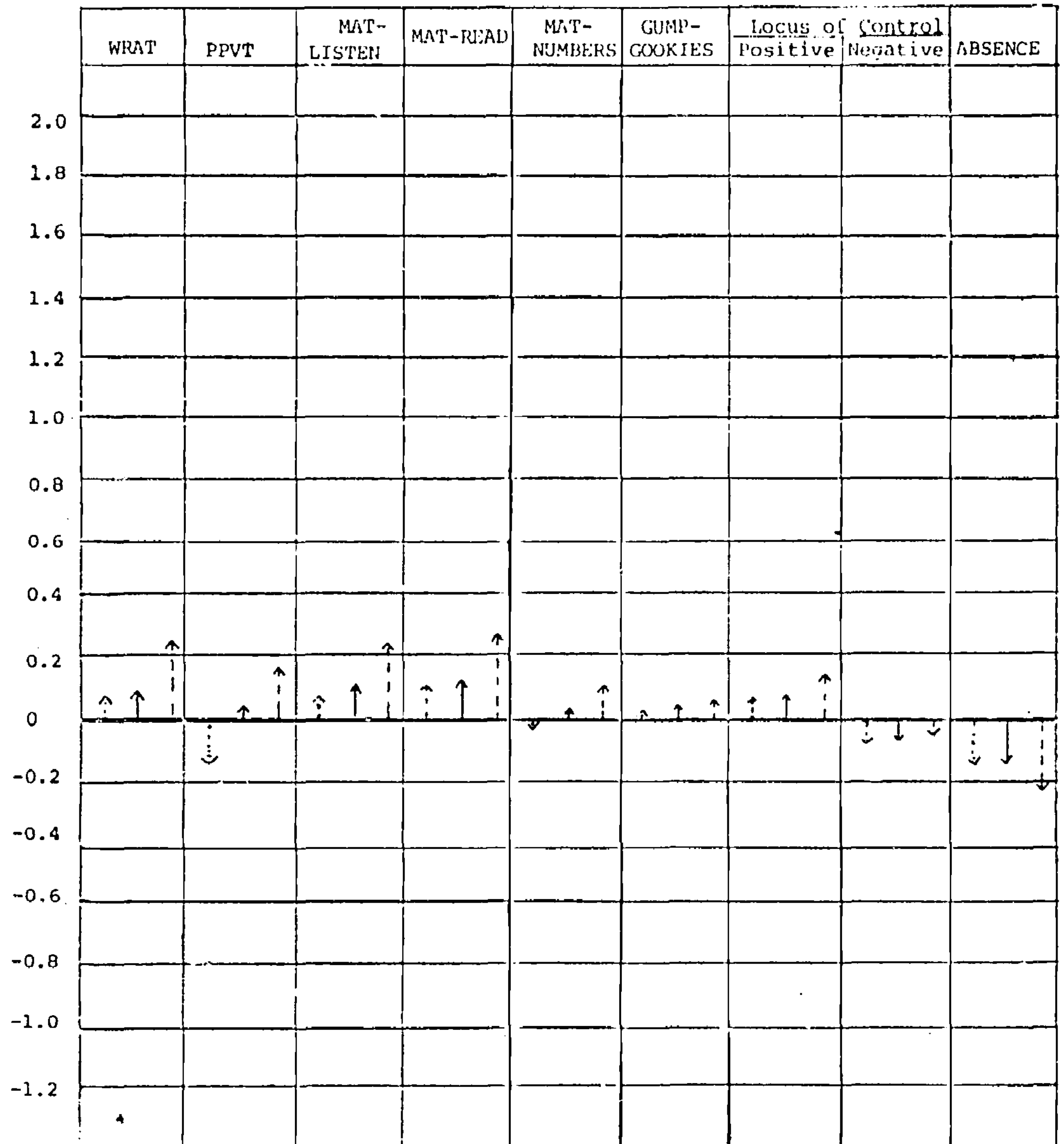
Figure VII - 22

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 2 *

Cohort III,
Kindergarten

N = 605





		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-GOOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	.726	-1.000	.733	.609	-.053	.036	.174	-.121	-1.712
	Adjusted	1.072	.310	.872	.722	.080	.100	.196	-.113	-1.742
	Unadjusted	2.942	1.384	1.639	1.325	.691	.164	.319	-.070	-2.717
*B/S = Effect in Standard Deviations	True Score Adj.	.057	-.131	.094	.108	-.008	.004	.086	-.075	-.144
	Adjusted	.084	.041	.112	.128	.012	.012	.097	-.070	-.147
	Unadjusted	.231	.182	.211	.236	.104	.020	.158	-.043	-.229
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

test. In addition, at this level the Far West Lab program appears to be having success, in at least some schools, in developing reading and arithmetic achievement, as measured by the MAT. While positive affective results were not found at either the class or child levels of analysis, positive achievement trends were found at the class level, and, to a lesser extent, at the child level.

The fact that the program's effects were found to differ with the level of analysis employed may, of course, reflect differences in the way in which the variables interrelate at the various levels of aggregation. More importantly, perhaps, these findings suggest that the Far West Lab program may be having a differential impact on the types of children and families served, or on the types of classes and/or schools in which the program is implemented.

Although we found similarities in the demographic characteristics of the children and families served by this Sponsor at all three levels of analysis, the analytic groups were not identical. It is possible that the positive achievement trends at the class and school levels of analysis reflect the particular makeup of the children included in these groups. Given a slightly different sample, at the child level of analysis, trends are less clear. Furthermore, while the teacher data suggest that Far West Lab teachers, in general, value the goals and practices advocated by the Sponsor, given the flexible nature of the Far West program, it is likely that a great deal of variability in actual program delivery is taking place. We have not yet merged either teacher self-report or observation data with child outcomes. Future analyses will explore the impact of the Far West program on varying types of children, classes, and schools in an attempt to identify the context(s) in which this program works best.

4.2.0 SPONSOR 3: UNIVERSITY OF ARIZONA, TUSCON EARLY EDUCATIONAL MODEL

The University of Arizona program focuses upon four general areas of developmant: language, reasoning abilities, motivation, and social arts and skills. The classroom environment is designed to reflect the child's home and community environment, so that skills and concepts may be learned in a natural, functional setting. One-to-one adult-child interactions and small group activities are utilized to individualize instruction and to help develop effective social interaction and communication skills.

4.2.1 School Level FT/NFT Contrasts

The subset of schools included in the analysis for this Sponsor was drawn from six sites -- two in the Northeast, three in the North Central region of the United States, and one in the South. Three of these sites are large cities, two are small cities, and one is medium sized. Approximately 43% of the FT schools and 40% of the NFT schools are located in the two North Central large city sites.

First, we shall compare the University of Arizona FT schools to the total group of FT schools for all Sponsors.

The FT schools for this Sponsor serve families of somewhat higher SES than the average FT school. Both mean adjusted income level and the mean percentage of mothers completing high school are above average for this group. So, too, the University of Arizona's FT schools serve children with slightly higher entering achievement levels than does the average FT school, as measured by scores on the Fall WRAT. The mean percentage of White pupils in the FT schools is approximately 53%, which is also somewhat high relative to the total FT group for all Sponsors.

Now we shall compare the FT/NFT schools for the University of Arizona program. Despite the relatively high SES and entering achievement levels of the FT group, the mismatch between the FT and NFT group is sizeable for the Arizona program. In fact, with the possible exception of Bank Street, this Sponsor's FT group starts out with the severest handicap in relation to its NFT group on both SES and entering achievement level. This is due to the fact that the University of

Arizona NFT group is higher than any other NFT group on adjusted income level, mothers' education, and Fall WRAT scores. The mean percentage of White pupils in the NFT schools is 69%, which is also higher than that for FT schools.

Figure VII-3 presents the school level FT/NFT contrasts for the University of Arizona on each of the outcome variables. There is only one significant school level contrast when initial differences are partialled out: the NFT group exceeds the FT group significantly on the MAT arithmetic subtest. In addition, there are several other trends in these data. The FT group exceeds the NFT group on the WRAT and Gumpgookies test; on the other hand, the NFT group exceeds the FT group on the MAT listening and Locus of Control (positive) subtests. Finally, the variability across schools on the other outcomes suggests that some FT schools may be having positive effects in these areas as well.

4.2.2 Class Level FT/NFT Contrasts

The subset of classes analyzed was drawn from the same sites included in the school analyses. However, the distribution of classes by site is somewhat different from the distribution of schools. The large North Central sites account for less of the total group at the class level -- approximately one-third -- and the distribution of the remaining classes by site is also somewhat different.

At the class level of analysis, there are still large differences between the FT and NFT groups in SES and entering achievement level. These differences once again favor the NFT group. The mean percentage of White children in FT classes (48%) is somewhat lower than the mean percentage for NFT classes (67%), which also parallels the school level.

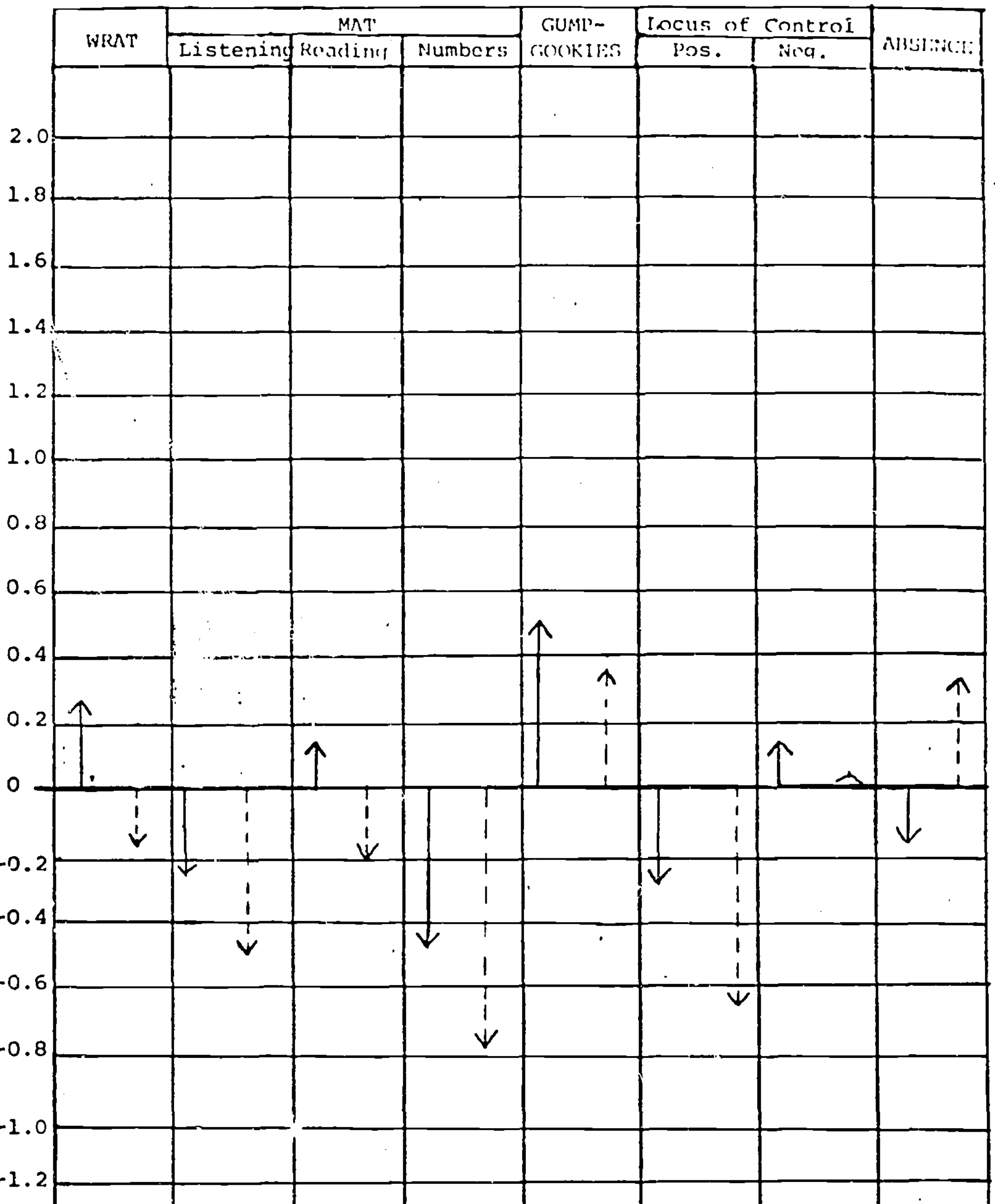
Figure VII-23 presents the FT/NFT contrasts for the Spring outcome measures. Despite the initial advantage of the NFT group, covariance adjustment benefits this group at the class level of aggregation. In Chapter VIII we will explore in greater detail the relationship of entry level and post-test scores, with implications for the effects of covariance adjustment. For the moment, it is important to note that this reversal in the way in which adjustment works for this Sponsor raises serious questions about the validity of the class level contrasts. Even with these shifts in the nature of adjustment, however, the school

Figure VII - 3
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 3

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:
Cov. Adj. Effect
Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.239	-0.833	0.386	-1.463	5.649	-0.145	0.053	-0.754
	Unadj.	-0.149	-1.920	-0.478	-2.488	4.080	-0.351	0.005	1.678
SE = Standard Error of B	Adjusted	0.174	0.817	0.567	0.715	2.689	0.139	0.105	1.252
	Unadj.	0.246	1.006	0.698	0.864	2.949	0.169	0.103	1.465
t = B/SE = "Significance" Statistic	Adjusted	1.374	-1.020	0.681	-2.047	2.101	-1.045	0.504	-0.603
	Unadj.	-0.606	-1.908	-0.685	-2.879	1.384	-2.077	0.049	1.145
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.274	-0.223	0.159	-0.463	0.501	-0.265	0.160	-0.150
	Unadj.	-0.171	-0.513	-0.196	-0.788	0.362	-0.641	0.015	0.335
N = Number of Schools in Computation	FT	21	21	21	21	21	21	21	21
	NFT	20	20	20	20	20	20	20	20

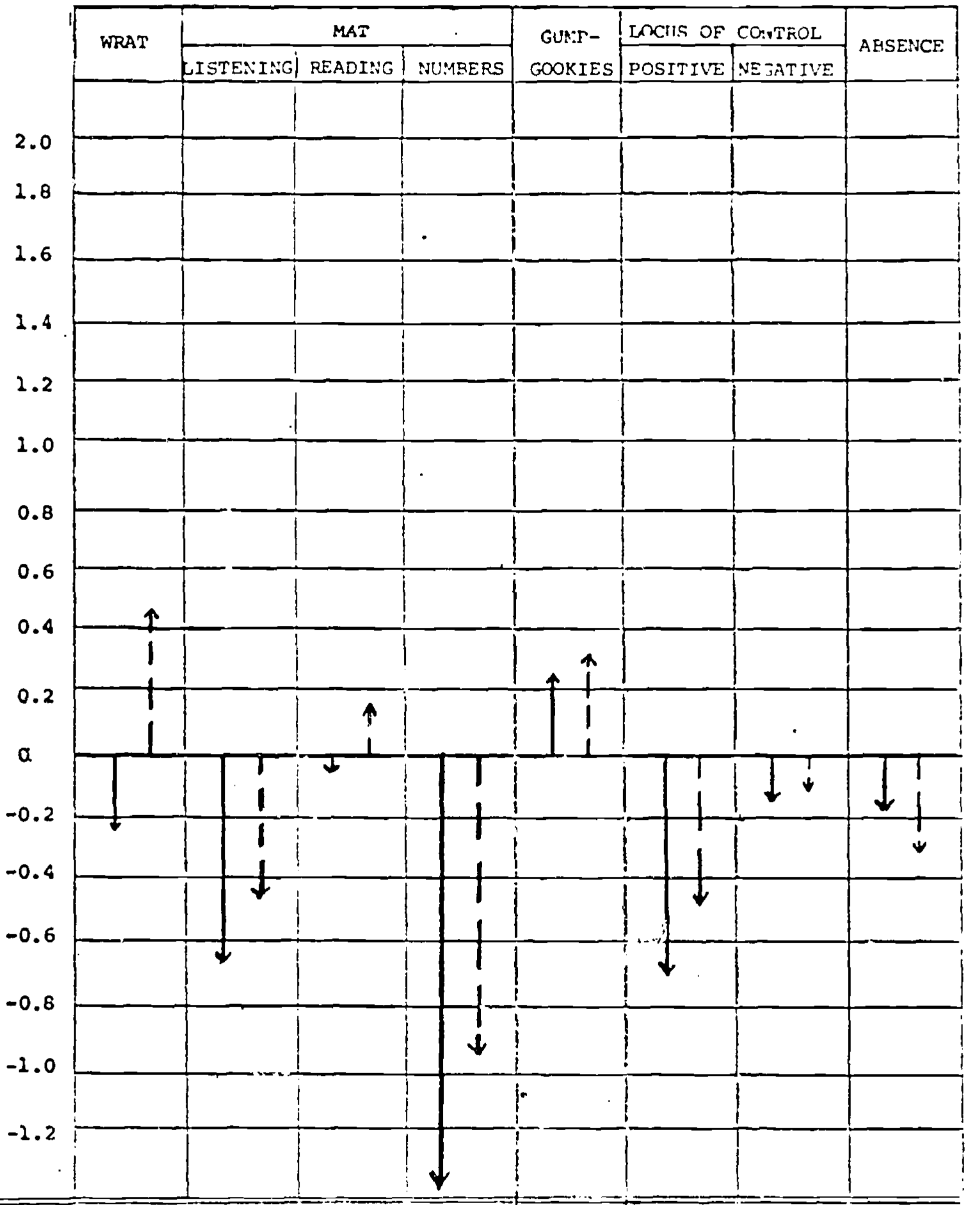
Figure VII - 23

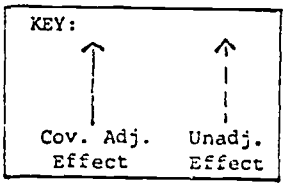
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 3*

Cohort III,
Kindergarten

N = 64





		WRAT	LISTENING	READING	NUMBERS	GUMP-GOOKIES	POSITIVE	NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	-1.466	-2.813	-.030	-4.842	.721	-.546	-.080	-1.186
	Unadj.	3.069	-1.852	.503	-3.495	1.055	-.367	-.056	-1.624
* B/S = Effect in Standard Deviations	Adjusted	-.205	-.644	-.010	-1.295	.212	-.686	-.144	-.196
	Unadj.	.429	-.424	.166	-.935	.310	-.461	-.102	-.268
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

and class level profiles are remarkably similar. Only on the WRAT is there a significant change in the direction of the adjusted contrast. At the school level of analysis, the FT group exceeds the NFT group slightly on this variable; at the class level the reverse is true.

4.2.3 Child Level FT/NFT Contrasts

At the child level of analysis, the NFT group is similar in geographic distribution to the school level NFT group, whereas the FT group resembles the class level FT group in geographic distribution. That is, 32% of the FT group is located in two large, North Central sites; 42% of the NFT group is located in these sites.

At the child level of analysis, there are once again large differences between the FT and NFT groups. The NFT group is higher than the FT group on both SES and entering achievement level. Approximately 48% of the FT children and 75% of the NFT children are White, which also parallels school and class levels.

Figure VII-24 summarizes the FT/NFT contrasts on Spring measures at the child level of analysis. As in the school and class level analyses, the NFT group exceeds the FT group on the MAT listening and arithmetic subtests. However, there are no overall differences between the two groups on the other outcomes.

4.2.4 Selected Teacher Data

University of Arizona teachers are similar to the average FT teacher for all Sponsors in both age and experience. They have fewer advanced credits or degrees, however, than any other FT Sponsor group, and their salaries are relatively low. The NFT teachers for this Sponsor have many more years of teaching experience that the FT teachers. They also have more advanced credits and degrees than the FT group. Approximately 84% of the FT teachers and 88% of the NFT teachers are White.

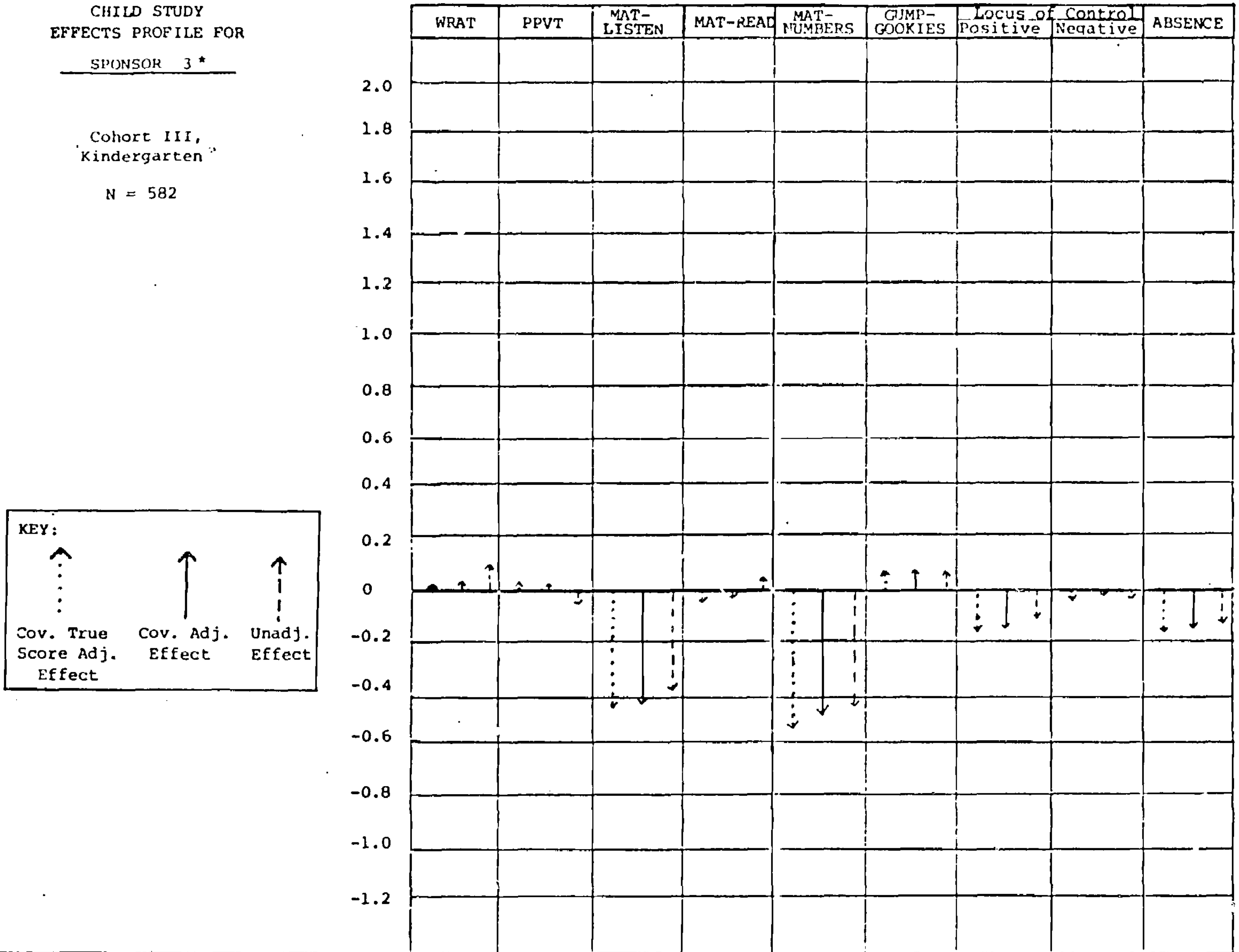
University of Arizona teachers report receiving more training in child-centered learning activities than the average FT teacher. They are more child-centered in their goals and reported practices than their NFT counterparts. They also make more visits to pupils' homes relative to their NFT counterparts and all other Sponsor groups.

Figure VII - 24
CHILD STUDY
EFFECTS PROFILE FOR

SPONSOR 3 *

Cohort III,
Kindergarten

N = 582



		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-GOOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	.091	.275	-3.343	-.161	-3.415	.641	-.313	-.021	1.911
	Adjusted	.356	.233	-3.237	-.074	-3.313	.690	-.297	-.015	1.888
	Unadjusted	1.367	-.379	-2.851	.240	-2.846	.618	-.255	-.023	1.781
*B/S = Effect in Standard Deviations	True Score Adj.	.007	.036	-.430	-.029	-.512	.076	-.155	-.013	.161
	Adjusted	.028	.031	-.417	-.013	-.496	.082	-.147	-.009	.159
	Unadjusted	.107	-.050	-.367	.043	-.426	.074	-.136	-.014	.150
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

4.2.5 Summary and Discussion

At the school level of analysis, the University of Arizona FT group exceeds the NFT group on the Gumpgookies test, a result which is paralled in the class level data. On the other hand, the NFT group for this Sponsor exceeds the FT group on the MAT listening and reading subtests, at all three levels of analysis. On each of the other Spring measures, the FT/NFT contrasts are extremely small, overall, or reverse direction from one level of analysis to another.

The negative achievement contrasts may reflect the severity of the mismatch between the two groups. The higher initial achievement of the NFT group may represent a faster rate of learning on the part of these children, a rate which might produce greater gains in achievement during the kindergarten year. This discrepancy may not influence pupil motivation as much as pupil achievement. Then too, the negative achievement contrasts may reflect the heavy emphasis of the Arizona program on the development of problem solving and social-emotional growth, an emphasis which is reflected in teachers' reported values. Whatever is producing these differences, they appear to be independent of differences in geographic location, which varied by level of analyses.

On the other hand, the lack of strong, consistent contrasts in the other areas suggests that in at least some areas of child growth and development there is variability in the effect the Arizona program has on children. This variability may be the result of geographic differences which may be associated with differences in teacher delivery. Like several of the other child-centered approaches, the Arizona program relies primarily on the teaching staff to plan the learning environment for children, develop appropriate learning activities, and provide appropriate responses to children's needs. While overall, the FT teachers report holding values that reflect the Arizona orientation, it is likely that there is a great deal of variability in the way in which teachers implement the model in the classroom. Future analyses are needed to determine whether or not there is, in fact, variability in program implementation, and whether or not these differences affect pupil outcomes.

4.3.0 SPONSOR 5: BANK STREET COLLEGE APPROACH

The Bank Street approach is designed to change the school system to meet the developmental needs of children. Heavy emphasis is placed on teacher training to help teachers organize the classroom environment for self-directed learning and to plan events to meet the needs of children. The individualized, flexible curriculum is designed not only to help children acquire basic skills, but also to help children master how to learn. Creativity and self expression are also important program goals.

4.3.1 School Level FT/NFT Contrasts

The subset of schools meeting criteria for inclusion in the school level analyses were drawn from five Northeastern sites. Approximately half the schools are in small towns, of between 10,000 and 50,000 population. The remainder are fairly equally divided between large and medium-sized cities.

The Bank Street FT schools are similar to the average FT school for all Sponsors in entering achievement level. The mean adjusted income level for these schools is slightly above the average for all Sponsors; however, the Bank Street schools are no different from the average FT school in the mean percentage of mothers completing high school. The overall mean percentage of minority pupils in these FT schools is approximately 54%; however, there is a great deal of variation across sites.

The FT/NFT schools for this Sponsor are not well matched. The NFT schools are higher in both SES and entering achievement than the FT schools. The NFT schools also have a higher mean percentage of White pupils (62%). In fact, if we examine the index of mismatch for Bank Street (for which see Table VII-4) we find that the average difference between the adjusted and unadjusted FT contrasts at the school level of analysis are greater for this Sponsor than for any other.

Figure VII-14 summarizes the school FT/NFT contrasts for Bank Street on each of the Spring outcome variables. When we adjust for

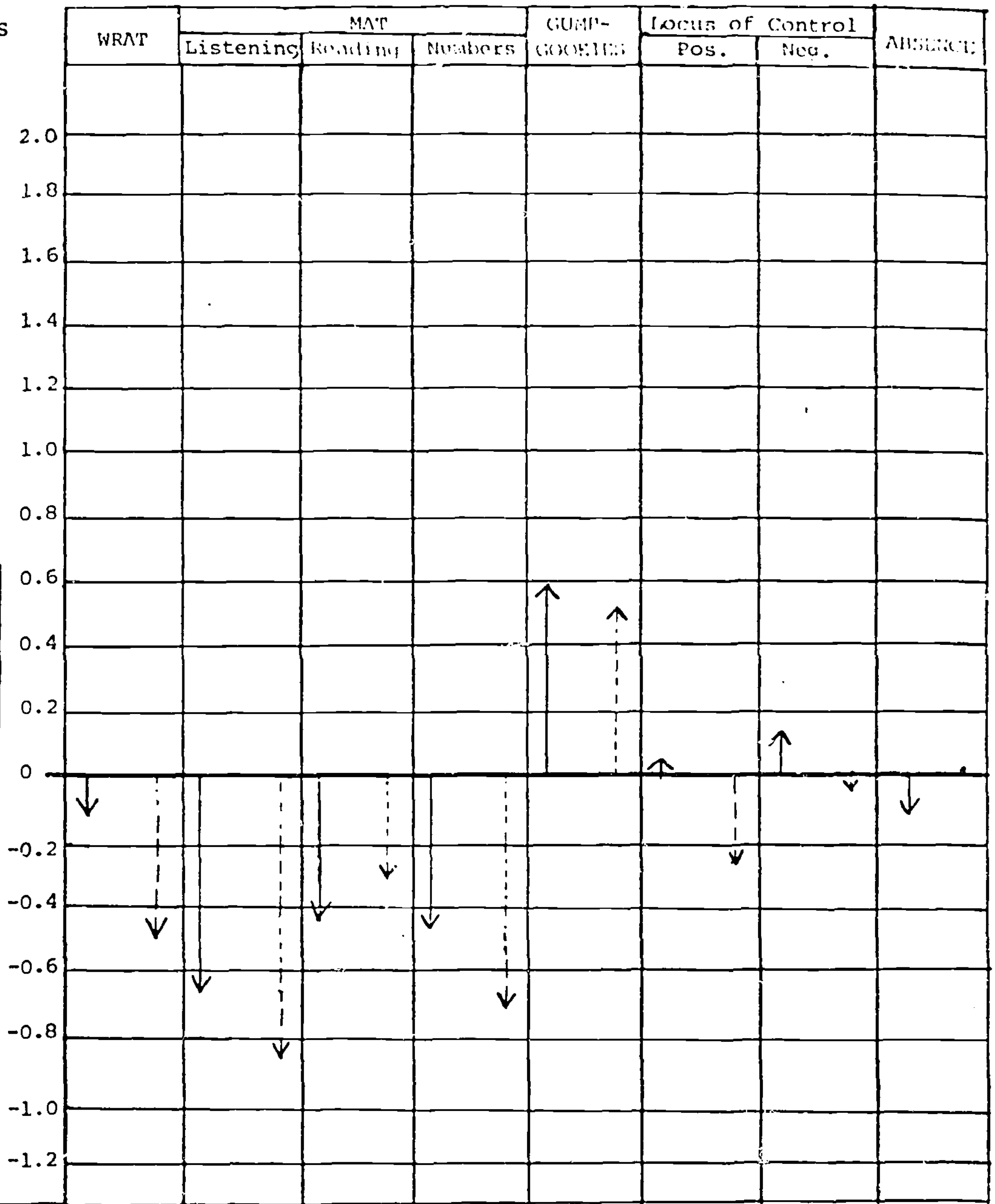
Figure VII - 14

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 5

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY: Cov. Adj. Effect — Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.094	-2.491	-1.022	-1.396	7.495	0.019	0.058	-0.899
	Unadj.	-0.452	-3.187	-1.839	-2.172	6.396	-0.139	-0.011	0.042
SE = Standard Error of B	Adjusted	0.204	0.959	0.651	0.793	3.639	0.164	0.127	1.696
	Unadj.	0.283	1.155	0.791	0.962	4.188	0.200	0.127	1.951
t = B/SE = "Significance" Statistic	Adjusted	-0.462	-2.597	-1.571	-1.760	2.060	0.118	0.459	-0.530
	Unadj.	-1.598	-2.759	-2.325	-2.253	1.527	-0.697	-0.085	0.021
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	-0.107	-0.669	-0.420	-0.450	0.598	0.034	0.165	-0.156
	Unadj.	-0.512	-0.856	-0 310	-0.700	0.510	-0.245	-0.031	0.007
N = Number of Schools in Computation	FT	16	16	16	16	16	16	16	16
	NFT	15	15	15	15	15	15	15	15

initial differences between the FT/NFT groups, the NFT group exceeds the FT group on all achievement measures. Although only one of these contrasts is significant (Listening to Sounds), the pattern is consistent across all the achievement outcomes.

On the other hand, there is one significant affective outcome that favors the FT group; the mean score of the FT schools on the Gumpgookies test exceeds the mean score of the NFT schools by .6 of a standard deviation. While there are no overall differences between the FT and NFT groups on any of the other affective outcomes, the variability of the Locus of Control contrasts across schools suggests that a certain number of FT schools may exceed their NFT counterparts on these measures as well.

4.3.2 Class Level FT/NFT Contrasts

The subset of classrooms included in the class level analysis is distributed somewhat differently from the subset of schools. While the FT classes are located in the same five sites, the proportion of classes in each geographic area is somewhat different. Whereas approximately half the FT and NFT schools were located in small, Northeastern cities, 45% of the FT classes and 75% of the NFT classes are located in these sites. On the other hand, while 30% of the FT classes are located in the large Northeastern cities, none of the NFT classes are located there.

Since the small city NFT group is predominantly White, and relatively high on adjusted income and mother's education level, the net result of this severe geographic mismatch is to magnify the demographic differences between the groups. Only with respect to the Fall WRAT are the differences between the two groups somewhat smaller at the class than at the school level of aggregation.

As seen in Figure VII-25, the FT group is exceeded by the NFT group on all Spring outcome variables at the class level of analysis. However, the fact that differences between the two groups on all outcomes are magnified, not decreased, by covariance adjustment, makes these contrasts extremely questionable. In fact, we shall find in the Entry Level Studies, Chapter VIII, that these results for Bank Street are spurious, due to

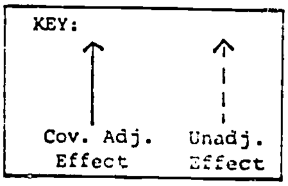
Figure VII - 25

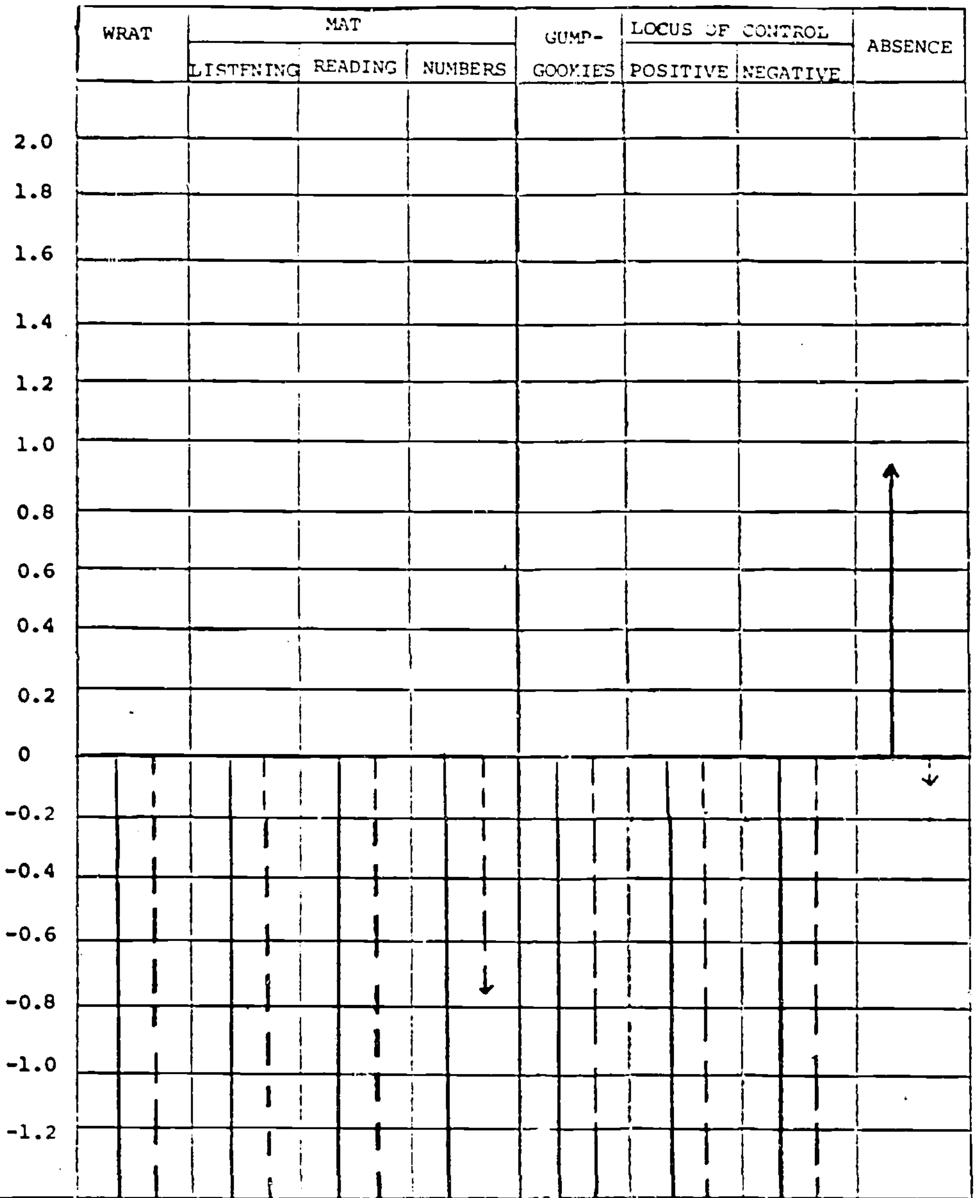
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 5*

Cohort III,
Kindergarten

N = 36





		WRAT	MAT LISTENING	MAT READING	MAT NUMBERS	GUMP-GOOKIES	LOCUS OF CONTROL POSITIVE	LOCUS OF CONTROL NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	-24.891	-17.512	-16.595	-8.053	-9.045	-3.322	-1.062	5.719
	Unadj.	-14.023	-12.965	-12.938	-2.841	-5.492	-2.246	-1.113	-.422
*B/S = Effect in Standard Deviations	Adjusted	-3.480	-4.008	-5.481	-2.154	-2.661	-4.168	-1.925	.945
	Unadj.	-1.961	-2.967	-4.273	-.760	-1.615	-2.818	-2.019	-.070
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

biases in sampling at the class level. These contrasts will not, therefore, be discussed further here.

4.3.3 Child Level FT/NFT Contrasts

The subset of FT/NFT children meeting criteria for inclusion in the child level analysis are distributed differently across sites than either schools or classes. However, the NFT group still exceeds the FT group substantially on both SES and entering achievement levels.

Figure VII-26 presents the FT/NFT contrasts for Bank Street at the child level of analysis. The child achievement results parallel those obtained at the school level. There are shifts in the affective results, however. The FT group no longer exceeds the NFT group on the Gumpgookies test, but does exceed the NFT group on the Locus of Control (negative) and attendance measures. On the average, the FT pupils are absent seven fewer days than the NFT pupils.

4.3.4 Selected Teacher Data

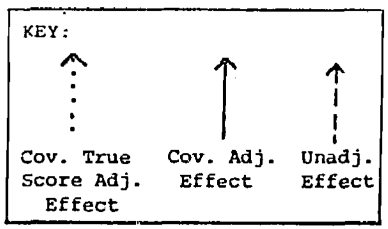
Bank Street teachers are similar in age to FT teachers in other programs. Although they have had slightly less experience overall than the average FT teacher, they have been in their present schools an average number of years and in the Sponsor's approach somewhat longer than average. Bank Street FT teachers are the most highly educated teacher group; approximately 95% have obtained advanced credits or degrees. They are also among the highest paid. Approximately 78% of the Bank Street FT teachers are White. The NFT teachers for this Sponsor are somewhat older and more experienced than the FT teachers, on the average. However, they are not as highly educated. Approximately 71% of the Bank Street NFT teachers are White.

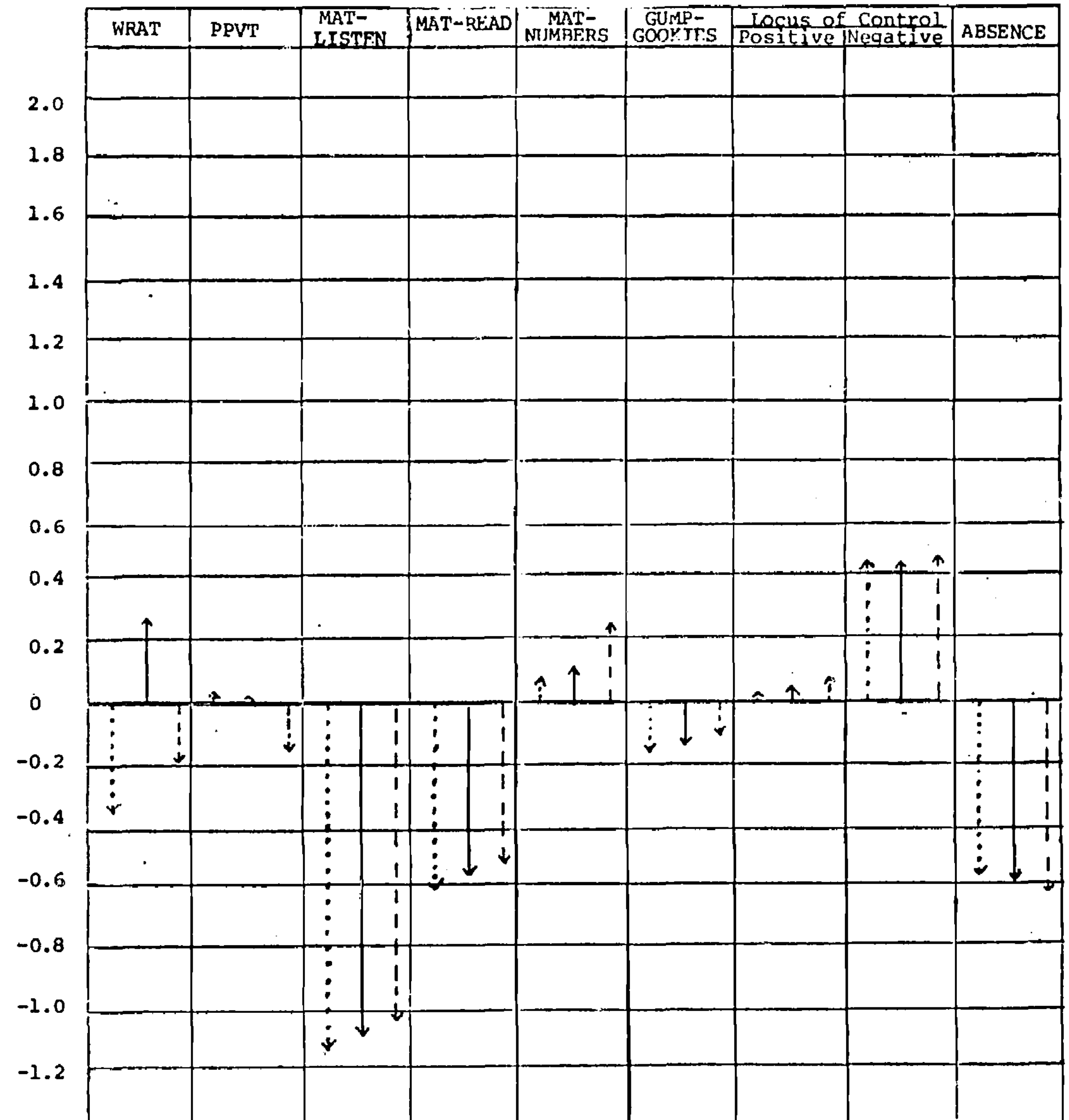
Bank Street teachers report receiving relatively little training, overall. What training they do report is in the area of child-centered philosophy and practices. Whether these reports reflect reality or heightened expectations is difficult to say. However, the FT teachers do report valuing social skills development and child-centered goals and practices.

Figure VII - 26
CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 5 *

Cohort III,
Kindergarten

N = 343





		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-GOOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	-4.354	.257	-8.685	-3.375	.629	-1.246	.025	.740	-6.991
	Adjusted	3.771	.029	-8.451	-3.185	.852	-1.137	.061	.754	-7.042
	Unadjusted	-2.475	-1.191	-8.013	-2.866	1.628	-.972	.153	.769	-7.350
*B/S = Effect in Standard Deviations	True Score Adj.	-.341	.034	-1.118	-.601	.094	-.149	.012	.459	-.589
	Adjusted	.296	.004	-1.088	-.567	.128	-.136	.030	.459	-.593
	Unadjusted	-.194	-.156	-1.031	-.510	.244	-.116	.076	.477	-.619
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

4.3.5 Summary and Discussion

The severe mismatch between the FT/NFT groups may account for the fact that the NFT group exceeds the FT group on each of the achievement outcomes at the school level of analysis. While we have found that fan spread may not be operating consistently across all Sponsors, it is quite likely that it is in effect in this particular case. The home background factors that led to the initial advantage of the NFT group over the FT group for this Sponsor are probably continuing to affect the learning rate of the NFT children. Although covariance adjustment may control for the initial differences between the two groups, it may not be able to adjust adequately for differential learning rates.

Despite the failure to find positive achievement contrasts, Bank Street does appear to be having some success in achieving its affective goals at the school level of analysis. Bank Street FT children appear to have more motivation to achieve and enjoy school more, as measured by the Gumpgookies test, than do their NFT counterparts. Also, while there are no overall differences between the FT and NFT schools on Locus of Control, Bank Street may be having an impact on at least some schools in this area as well.

The variability in affective findings from school to child level of analysis may be a function of at least two factors. (Since the class level findings are subject to statistical problems, they will not be discussed here.) First, the variability may be related to differences in the sample of children analyzed at the different levels. However, this possibility seems unlikely since the relative differences between the FT/NFT groups remains the same.

Second, the change in geographic distribution from school to class level of analysis may reflect important differences in program implementation associated with sites, schools, teachers, or parents. We shall find in Monograph III that Bank Street, like many other Sponsors, has had varied success in delivering its approach to its several sites. The Bank Street approach seeks to change teachers and school/community systems, not merely pupil test scores. Moreover, its individualized philosophical

orientation is inconsistent with prescribed teacher training techniques or structured curriculum materials. Thus, it is highly dependent upon the individual receptivity and competency of teachers, school administrators, and community persons for its successful implementation. We will examine these implementation questions in much greater detail in future reports.

Despite this variability across levels of analysis, it does appear that Bank Street is having some impact in the affective development of its pupils. Moreover, while Bank Street's FT teachers report valuing the child-centered approach highly, so do its NFT teachers. That is, both groups place less value on the development of basic skills in kindergarten than on the encouragement of exploration, manipulation, and self-confidence. In light of this, it appears that Bank Street may be delivering something to at least some teachers that assists them in achieving their objectives. It remains for future analysis to determine under what circumstances these objectives are achieved and whether or not these motivational advances are translated into achievement gains.

4.4.0 SPONSOR 7: UNIVERSITY OF OREGON

The University of Oregon approach is designed to teach children the basic skills of reading, arithmetic, and language. It is based on the assumption that disadvantaged children can perform at "normal" levels of achievement when the instructional program builds on the skills children bring with them to school. The curriculum materials are programmed and sequenced. Teaching techniques are highly prescribed with heavy e phasis on structured small group instruction, quick paced questic d-answer periods, and the use of positive reinforcement to shape avior.

4.4.1 School Level FT/NFT Contrasts

The subset of FT schools included in the analyses for this Sponsor was drawn from two middle-sized cities in the North Central region and one middle-sized city in the Northeast. The NFT schools were drawn from the same sites, with the addition of one NFT school in a large Northeastern city.

Compared to other Sponsors, the University of Oregon program has the poorest group of FT schools. The mean adjusted income for these schools is lower than for any other Sponsor group. The FT schools for this Sponsor are also slightly below average on mother's education. On the other hand, these schools are average in entering achievement levels, as measured by the Fall WRAT. They serve predominantly non-White children.

Compared to its NFT group, this FT group is also far lower on mean adjusted income and slightly lower on mother's education level. The two groups are well matched on entering achievement level and percent minority, however. The mean percentage of minority children for FT schools is 77% and it is 72% for NFT schools. Apart from the lower poverty level of the FT groups, the FT/NFT schools appear to be relatively well matched.

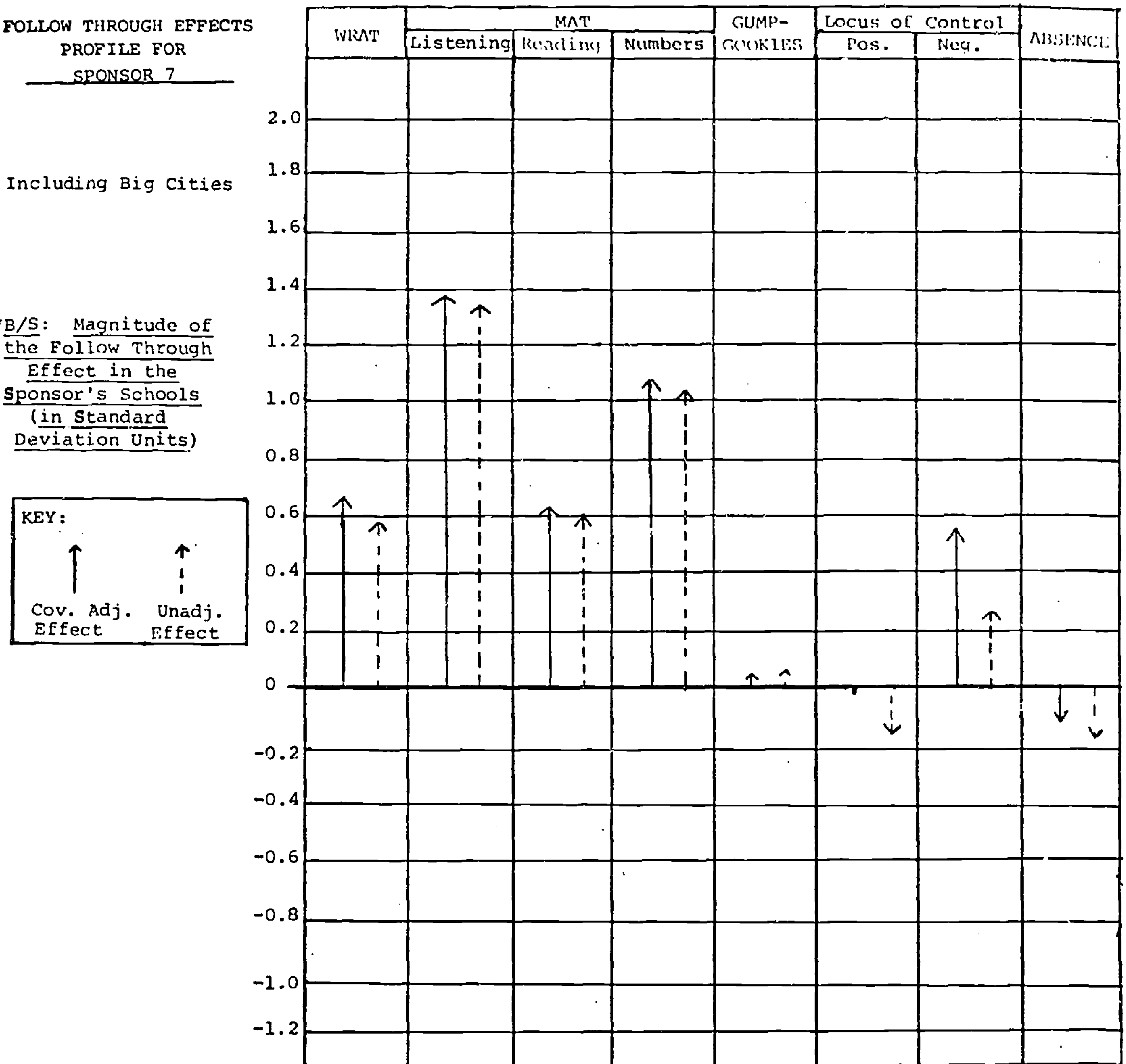
Figure VII-15 displays the FT/NFT contrasts for the Spring outcome measures for the University of Oregon program. With initial

Figure VII - 15.

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 7

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.595	5.202	1.503	3.308	0.560	-0.0003	0.195	-0.819
	Unadj.	0.528	5.114	1.460	3.199	0.621	-0.092	0.097	-0.978
SE = Standard Error of B	Adjusted	0.251	1.181	0.801	0.976	4.479	0.202	0.156	2.087
	Unadj.	0.344	1.404	0.962	1.170	5.091	0.243	0.154	2.372
t = B/SE = "Significance" Statistic	Adjusted	2.371	4.405	1.877	3.388	0.125	-0.002	1.250	-0.392
	Unadj.	1.536	3.643	1.518	2.734	0.122	-0.377	0.630	-0.412
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.673	1.397	0.618	1.066	0.045	-0.0005	0.554	-0.140
	Unadj.	0.598	1.374	0.600	1.031	0.050	-0.161	0.274	-0.170
N = Number of Schools in Computation	FT	11	11	11	11	11	11	11	11
	NFT	10	10	10	10	10	10	10	10

differences partialled out, the FT schools exceed the NFT schools on all achievement outcomes, and three of these four contrasts are statistically significant. There is also a trend for the FT group to exceed the NFT group on the Locus of Control (negative) measure. While there are no mean differences on the other outcomes measured, there is a great deal of variability across schools.

4.4.2 Class Level FT/NFT Contrasts

The FT/NFT classes included in these analyses are similar in geographic distribution to the schools described above. They are primarily in middle-sized cities in the Northeast and North Central regions. However, three FT classes are located in the large Northeastern city and none of the NFT classes are located there.

These three FT classes, which comprise approximately 14% of the FT group, are higher in SES and entering achievement than the other FT classes in the Oregon program. Despite the addition of these classes, the differences between the FT/NFT groups at the class level parallel those at the school level. The FT group is lower in adjusted income level and mother's education but equal or slightly above in entering achievement level. The mean percentage of non-White pupils in FT classes is 81% and for NFT classes it is 68%.

Figure VII-27 displays the FT/NFT contrasts for the University of Oregon at the class level of analysis. The profile is very similar to that found at the school level. The FT group exceeds the NFT group on all achievement outcomes as well as on the Locus of Control (negative) measure. In addition, the FT group exceeds the NFT group in attendance at this level of analysis. On the average children in FT classes are absent three days less than children in NFT classes.

4.4.3 Child Level FT/NFT Contrasts

Virtually all the FT/NFT children included in these analyses are located in middle-sized cities in the Northeast and North Central region. As in the school and class subsets, the FT children are lower in SES, and similar or slightly above the NFT children in entering

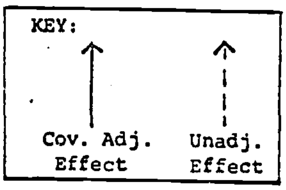
Figure VII - 27

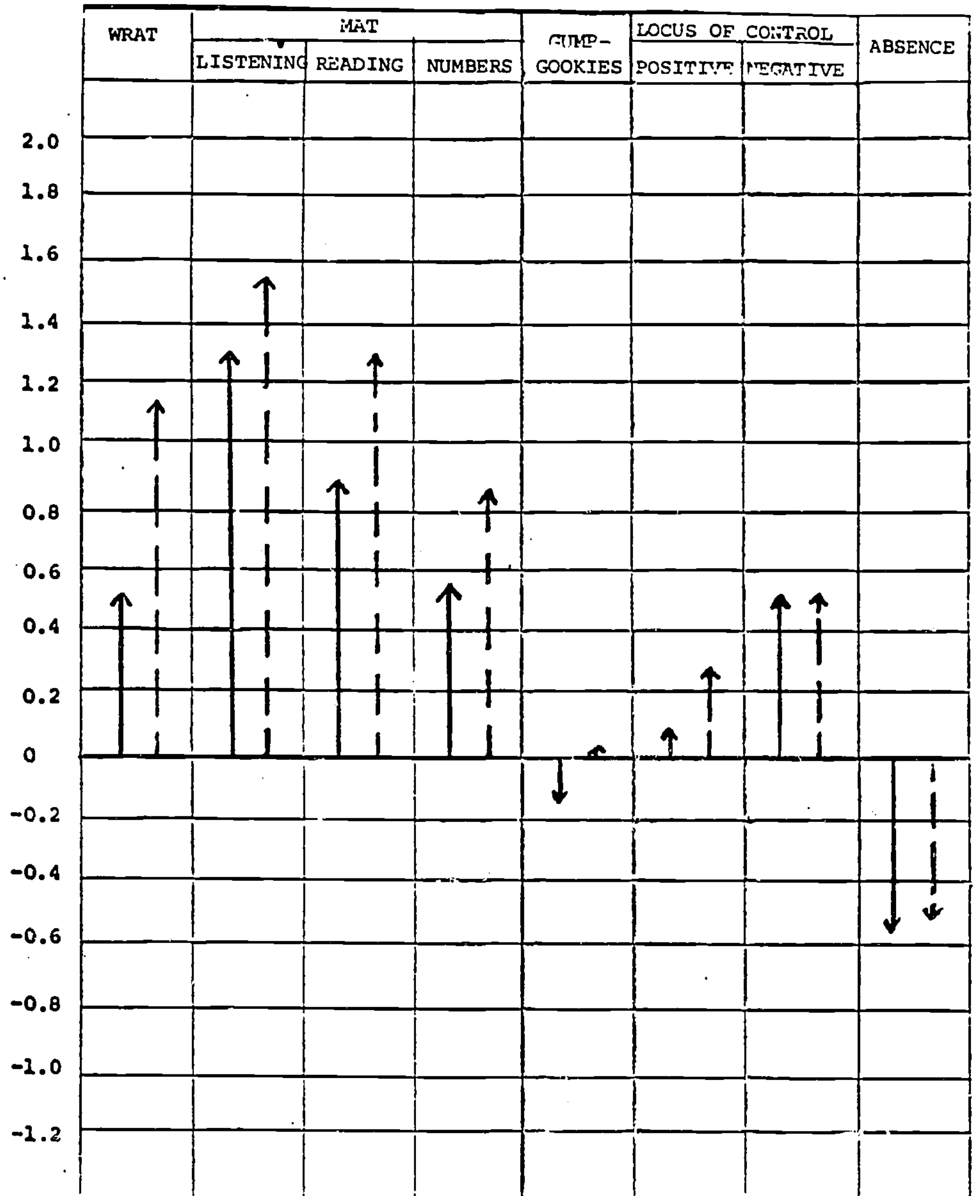
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 7 *

Cohort III,
Kindergarten

N = 33





		WRAT	MAT LISTENING	MAT READING	MAT NUMBERS	GUMP-GOOKIES	LOCUS OF CONTROL POSITIVE	LOCUS OF CONTROL NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	3.828	5.803	2.678	2.192	-.343	.076	.298	-3.548
	Unadj.	7.901	6.773	3.885	3.153	.013	.236	.290	-3.069
*B/S = Effect in Standard Deviations	Adjusted	.535	1.328	.885	.586	-.101	.096	.540	-.587
	Unadj.	1.105	1.550	1.283	.843	.004	.296	.526	-.507
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

achievement level. Approximately 79% of the FT children and 74% of the NFT children are non-White.

The FT/NFT child level contrasts are displayed in Figure VII-28. As can be seen, the FT group exceeds the NFT group on all achievement measures at this level of analyses as well. However, there are no significant child level contrasts on any of the affective outcomes or on the PPVT.

4.4.4 Selected Teacher Data

The University of Oregon FT teachers are younger than any other Sponsor FT group. They also have less experience and lower salaries than any other FT Sponsor group. Approximately 78% of these FT teachers have obtained graduate credits or degrees, which is similar to the overall educational attainment for all FT teachers. Finally, the University of Oregon program has more minority teachers (44%) than any other FT group.

The NFT teachers for this Sponsor are older, more experienced, and more highly educated than the FT teachers. Only 16% of the NFT teachers are from minority groups.

FT teachers in the Oregon program report receiving a great deal of training in structured learning activities, but little in other areas. They place greater value on the development of basic skills and the use of a structured learning environment than their NFT counterparts and other FT teachers as well. On the other hand, they place less value on the development of social skills or on involving parents in the school program compared to both their NFT counterparts and other FT teachers. They also make fewer home visits relative to their NFT group and other Sponsor FT groups.

4.4.5 Summary and Discussion

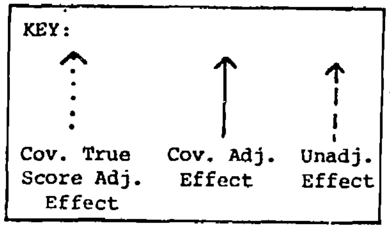
The University of Oregon program appears to be having a positive impact on achievement at all three levels of analysis. Given the low SES of these groups relative to other FT groups, the fact that the Oregon approach is having an impact on pupil achievement is encouraging

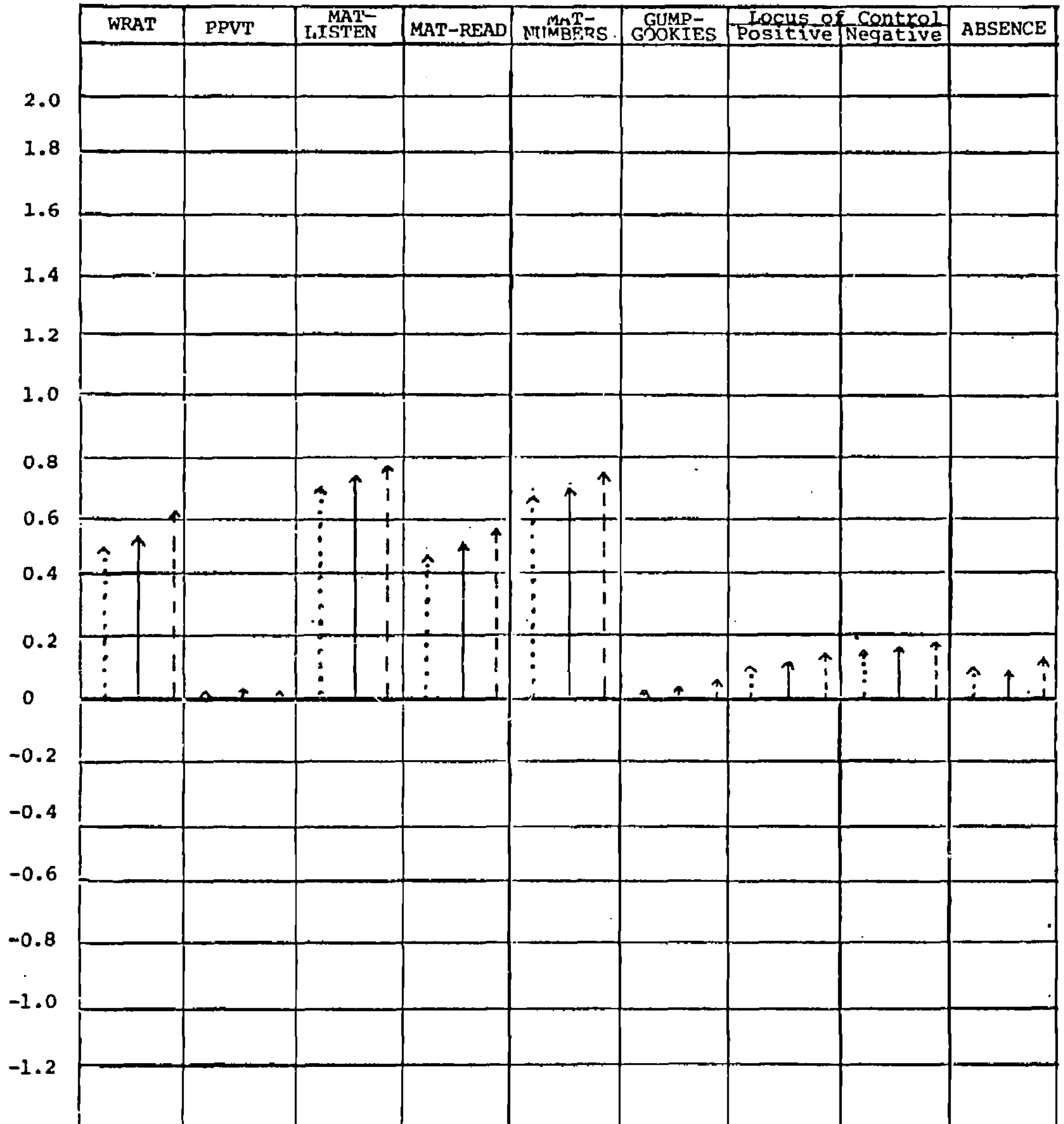
Figure VII - 28

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 7*

Cohort III,
Kindergarten

N = 272





		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-COOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	6.478	.051	5.551	2.656	4.557	.143	.253	.256	1.262
	Adjusted	7.013	.113	5.766	2.831	4.762	.244	.287	.268	1.215
	Unadjusted	8.107	.089	6.011	3.160	5.099	.451	.317	.282	1.375
*B/S = Effect in Standard Deviations	True Score Adj.	.508	.007	.714	.473	.683	.017	.126	.159	.106
	Adjusted	.550	.015	.742	.504	.713	.029	.142	.166	.102
	Unadjusted	.636	.012	.774	.562	.754	.054	.157	.175	.116
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

and in keeping with the program's objectives. The highly programmed curriculum materials and prescribed teaching materials may make this program's achievement effects less susceptible to variability in the children, classes, schools or communities served. This inference is somewhat premature, however, since the groups analyzed at all three levels were similar in community location and demographic characteristics.

The Oregon program has weak and variable effects in the affective domain, however. The FT groups do not exceed their NFT counterparts on achievement motivation. Nor is there consistency in the impact of the Oregon program on locus of control or attendance patterns.

Future analyses are needed to determine the effectiveness of this program in this area with different types of children, in different settings and over time.

4.5.0 SPONSOR 8: UNIVERSITY OF KANSAS BEHAVIOR ANALYSIS APPROACH

The primary objective of the University of Kansas approach is to facilitate the child's mastery of basic skills, particularly in reading and arithmetic through the establishment of a "token economy" within the classroom. Based on basic principles of behavior modification, the token exchange system is designed to provide systematic, positive reinforcement for desired behavior. The tokens, which are given as rewards for successful completion of tasks, may later be exchanged for desired activities. Within the "token economy" environment, programmed instructional materials are used to teach basic skills.

4.5.1 School Level FT/NFT Contrasts

The University of Kansas schools included in these analyses were drawn from six sites located in the Northeast, North Central, and Southern areas. Four of the sites are large cities, one a medium-sized city, and one a rural community. Over 50% of the FT/NFT schools are in the Northeast.

Compared to the total group of schools for all Sponsors, the University of Kansas schools are slightly below average on both indices of SES--mean percentage of mother's completing high school and mean adjusted income--and on entering achievement level. The Kansas schools also have more minority pupils than any other Sponsor group.

The FT/NFT schools are similar in ethnic composition. The mean percentage of non-White pupils in FT schools is 92%, in NFT schools it is 90%. The FT schools for this Sponsor are lower than the NFT schools in mean adjusted income. On the other hand, the FT schools exceed the NFT schools on mother's education. Finally, they exceed the NFT schools substantially on entering achievement, as measured by the Fall WRAT.

Overall, the difference between the FT/NFT schools is sizeable for this Sponsor. In fact, if we examine the index of mismatch (see Table VII-4) we find that the University of Kansas program has the largest mismatch in which the FT group exceeds the NFT group.

Whether these differences represent (1) Sponsor/district criteria for selecting FT schools, (2) a predisposition of active, relatively well educated parent groups toward this Sponsor, or (3) early treatment effects, they should be considered in examining the Spring contrasts.

Figure VII-16 presents the FT/NFT contrasts for the Kansas program on each of the Spring outcome measures. With initial differences partialled out, the FT group exceeds the NFT group on all achievement outcomes and on the Gumpgookies test (and these differences are statistically significant). On the other hand, there is a trend for the NFT group to exceed the FT group on the Locus of Control measures. There is no difference between the two groups on the absence measure.

4.5.2 Class Level FT/NFT Contrasts

The subset of classes included in the class level analyses are drawn from the same sites as the school level. A relatively greater proportion of FT and NFT classes, however are located in the large cities, and a smaller proportion in the medium-sized North Central site. Then too, there are some shifts in the ratio of FT to NFT schools in the various sites.

These changes in the distribution of FT/NFT classes do not change substantially the pattern of demographic characteristics found at the school level. The FT classes exceed the NFT classes on both mother's education level and initial achievement, differences which parallel the school data. Furthermore, the two groups are still predominantly non-White. However, the FT group is closer to the NFT group in adjusted income level at the class level than at the school level.

The class level FT/NFT contrasts are displayed in Figure VII- 29. The FT classes exceed the NFT classes on all achievement outcomes. There is also a trend for the FT group to exceed the NFT group on the Gumpgookies test. These results parallel the school level contrasts. On the other hand, the NFT group does not exceed the FT group on the locus of control measures at the class level.

Figure VII - 16

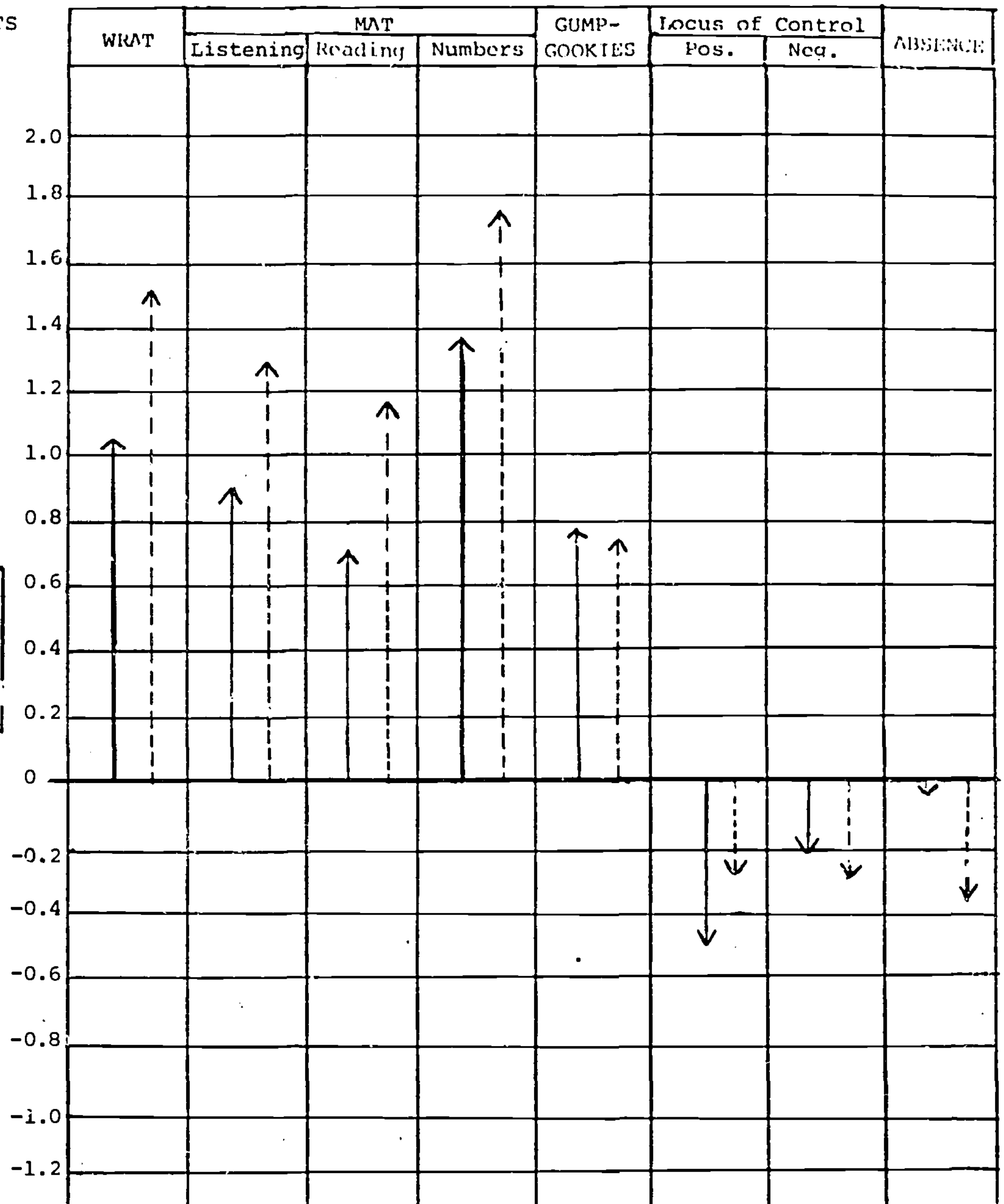
FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 8

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect — Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.934	3.416	1.749	4.333	9.934	-0.281	-0.071	-0.210
	Unadj.	1.372	4.808	2.876	5.478	9.515	-0.143	-0.092	-2.085
SE = Standard Error of B	Adjusted	0.195	0.919	0.624	0.760	3.487	0.157	0.122	1.625
	Unadj.	0.269	1.097	0.752	0.915	3.980	0.190	0.120	1.855
t = B/SE = "Significance" Statistic	Adjusted	4.790	3.716	2.804	5.700	2.849	-1.790	-0.586	-0.129
	Unadj.	5.101	4.382	3.825	5.987	2.391	-0.752	-0.763	-1.124
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	1.057	0.918	0.718	1.396	0.793	-0.495	-0.202	-0.036
	Unadj.	1.552	1.292	1.181	1.765	0.759	-0.251	-0.261	-0.362
N = Number of Schools in Computation	FT	20	20	20	20	20	20	20	20
	NFT	15	15	15	15	15	15	15	15

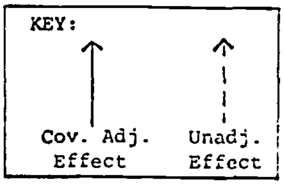
Figure VII - 29

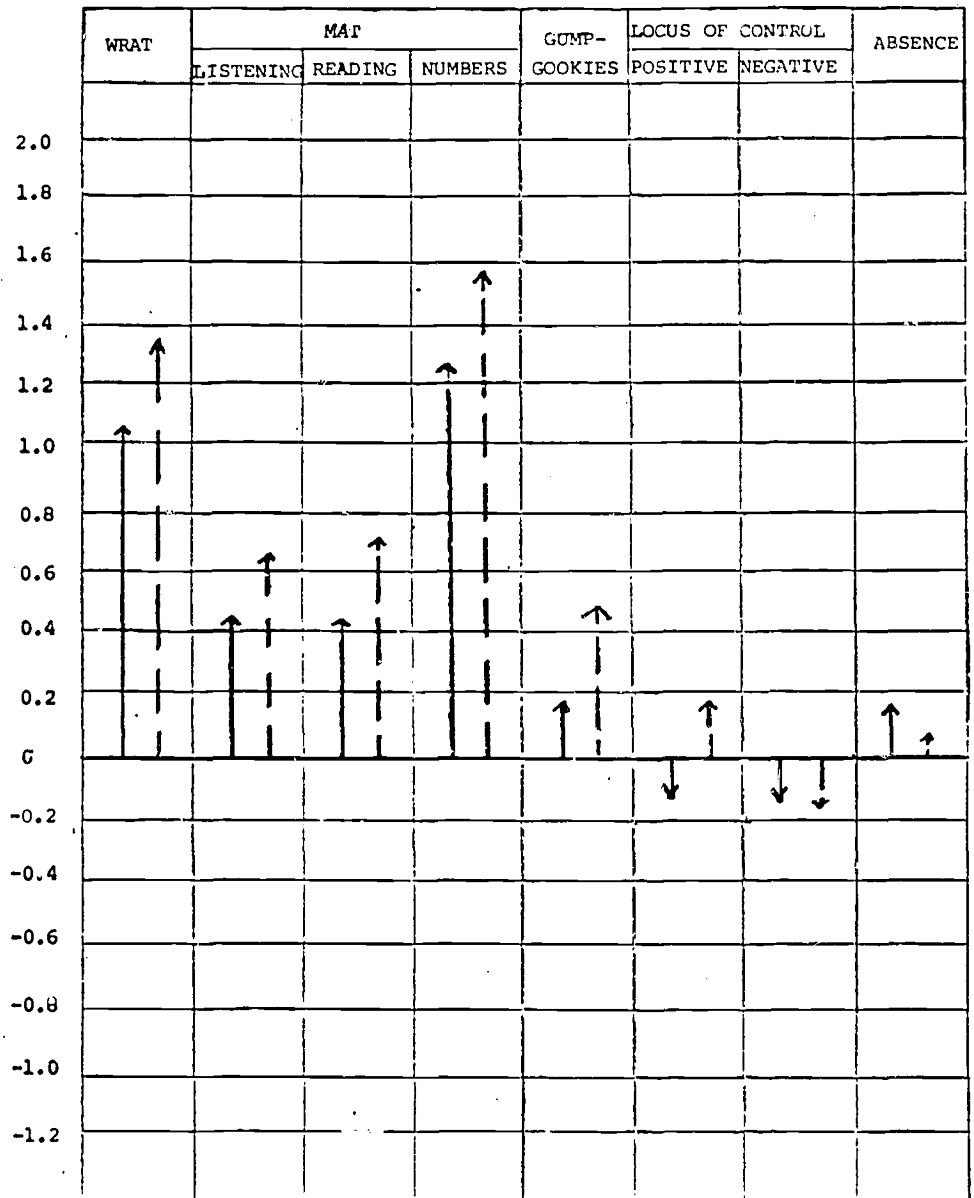
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 8*

Cohort III,
Kindergarten

N = 46





		WRAT	MAT LISTENING	MAT READING	MAT NUMBERS	GUMP-GOOKIES	LOCUS OF CONTROL POSITIVE	LOCUS OF CONTROL NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	7.285	1.804	1.265	4.727	.642	-.113	-.086	1.042
	Unadj.	9.826	2.816	2.155	5.904	1.826	.159	-.080	.393
*B/S = Effect in Standard Deviations	Adjusted	1.019	.413	.418	1.264	.189	-.141	-.156	.172
	Unadj.	1.374	.644	.712	1.579	.537	.199	-.146	.065
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

4.5.3 Child Level FT/NFT Contrasts

With small variations, the child level sample is similar to the class sample in geographic distribution.

The pattern of FT/NFT demographic differences for these children resembles that of the school level. The FT group is lower than the NFT group on adjusted income, higher on mother's education and entering achievement. Moreover, both groups are predominantly non-White.

It should be pointed out, however, that at the child level the FT group appears to be higher in initial achievement, relative to the total group for all Sponsors than at the school level. The two groups thus appear to serve different groups of children.

Figure VII-30 presents the FT/NFT contrasts for the child level analyses. At this level, the FT group exceeds the NFT group significantly on all achievement outcomes except the MAT listening for sounds subtest, where there is also an FT favoring trend. There are no important differences between the two groups on the PPVT or on any of the other outcome variables.

4.5.4 Selected Teacher Data

The FT teachers for the Kansas model are similar in age, experience, education, and salary to the total group of FT teachers for all Sponsors. There are slightly more minority teachers for this Sponsor (42%) than for the others. The NFT teachers, as a group, are similar to the FT teachers in their personal and professional background.

The University of Kansas teachers report receiving more training in the structured approach to teaching than any other Sponsor group. They also report receiving a great deal of training in working with parents and aides. Compared to their NFT counterparts, they place greater value on both these program components.

4.5.5 Summary and Discussion

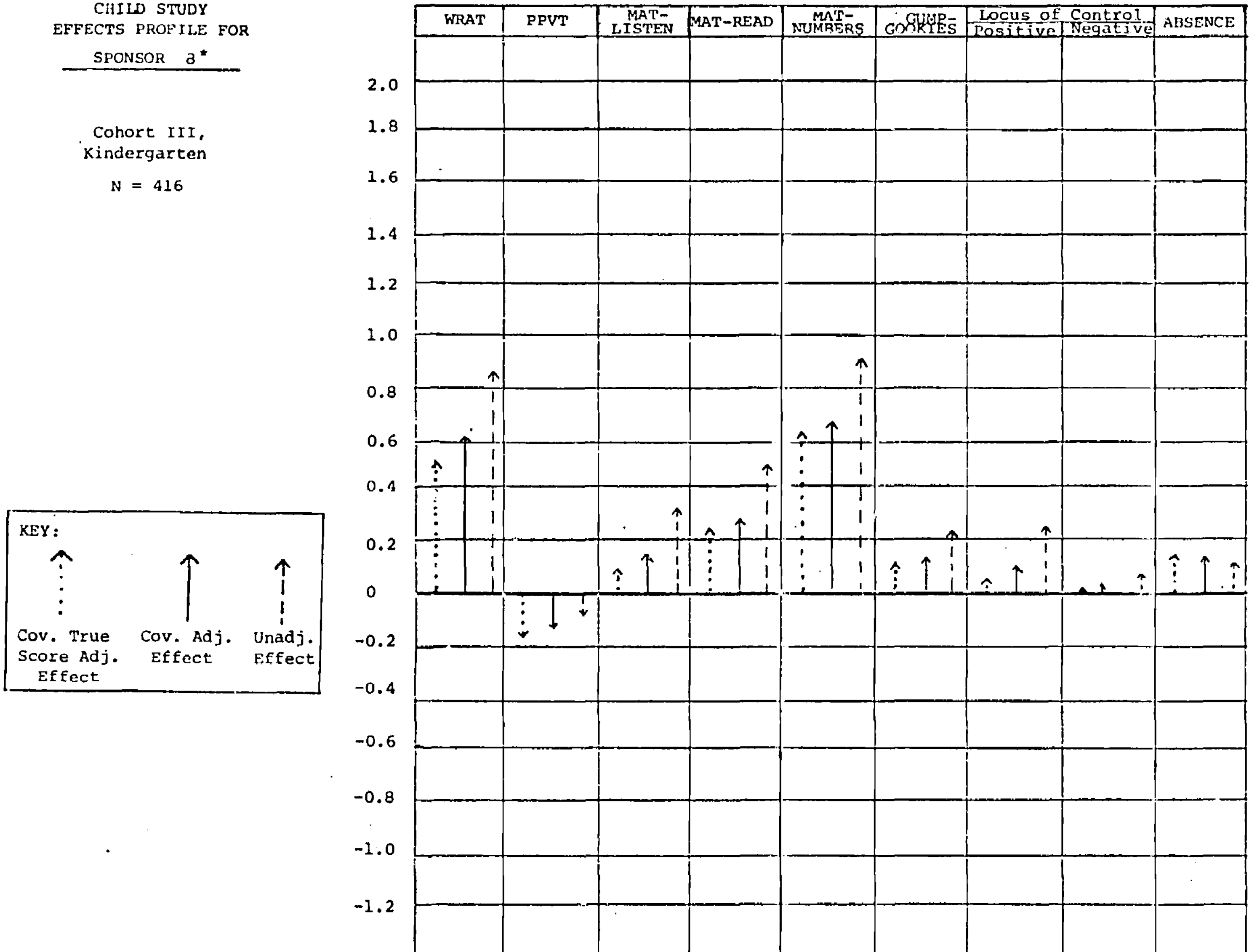
Despite small variations in geographical distribution and type of pupils included, the Kansas program appears to have strong positive

Figure VII - 30

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR a*

Cohort III,
Kindergarten

N = 416



		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	COOP-COOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	6.717	-1.353	.809	1.339	4.259	.994	.163	.036	1.864
	Adjusted	7.714	-1.199	1.209	1.664	4.642	1.181	.225	.059	1.776
	Unadjusted	11.233	-1.797	2.595	2.813	6.088	1.836	.451	.133	1.636
B/S = Effect in Standard Deviations	True Score Adj.	.527	-.178	.104	.238	.638	.119	.081	.022	.157
	Adjusted	.605	-.157	.156	.296	.695	.141	.112	.037	.150
	Unadjusted	.881	-.105	.334	.501	.912	.219	.224	.083	.138
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

achievement effects at all three levels of analysis.

It may be that the highly prescribed approach, with its strong emphasis on teacher training, produces consistent achievement results with a variety of types of children, in a variety of class, school, and community settings. However, we do not yet have sufficient data to draw this inference. Across all levels of analysis, and even within sites, the relationship between the FT groups and NFT groups for the Kansas program consistently favor the FT group on mother's education and initial achievement level. It may be that children coming from these better educated families not only come to school with higher achievement levels, but are more responsive to educational intervention than the other children. We have yet to see whether or not these contrasts emerge with a better matched NFT group.

The Kansas program also appears to have positive effects on achievement motivation, as measured by the Gumpgookies test. These effects vary, however, with the particular set of schools, classes, or children included in these analyses, suggesting that the motivation results are more influenced by the characteristics of the pupils served and the contexts in which the program operates. We will need to systematically explore the impact of the Kansas program on achievement motivation with various types of children, classes, schools, and communities.

Finally, we have not yet observed differences in the Kansas FT/NFT children on locus of control or attendance measures. In future years we will need to explore the development of these patterns over time.

4.6.0. SPONSOR 9: HIGH/SCOPE FOUNDATION, COGNITIVELY ORIENTED CURRICULUM

The High/Scope classroom environment may be described as "open," with an emphasis on active exploration, manipulation, and discovery. Within this open framework, the instructional approach is systematic and planned. The cognitively oriented curriculum is highly Piagetian. The ultimate goal is to develop in children the thinking skills they will need throughout their school years and adult lives.

4.6.1 School Level FT/NFT Contrasts

The subset of High/Scope schools included in these analyses was drawn from six sites: four large cities in the Northeast, North Central, and Western regions, and two small cities in the South and West. Over 63% of the FT schools and 54% of the NFT schools are located in the Western sites.

Compared to the total group of FT schools for all Sponsors, the High/Scope schools are below average in the mean adjusted income of the families served. They are average, however, in the mean percentage of mothers completing high school and in entering achievement level, as measured by Fall WRAT scores. The FT schools serve predominantly minority children.

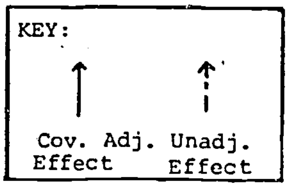
The High/Scope FT schools are lower in mean adjusted income than the NFT schools. Moreover, they have far more children from minority groups than the NFT schools. The mean percentage of minority children for FT schools is 81%; for NFT schools it is 60%. The two groups are similar in mother's education and entering achievement levels, however, so that overall there is a relatively close match between groups.

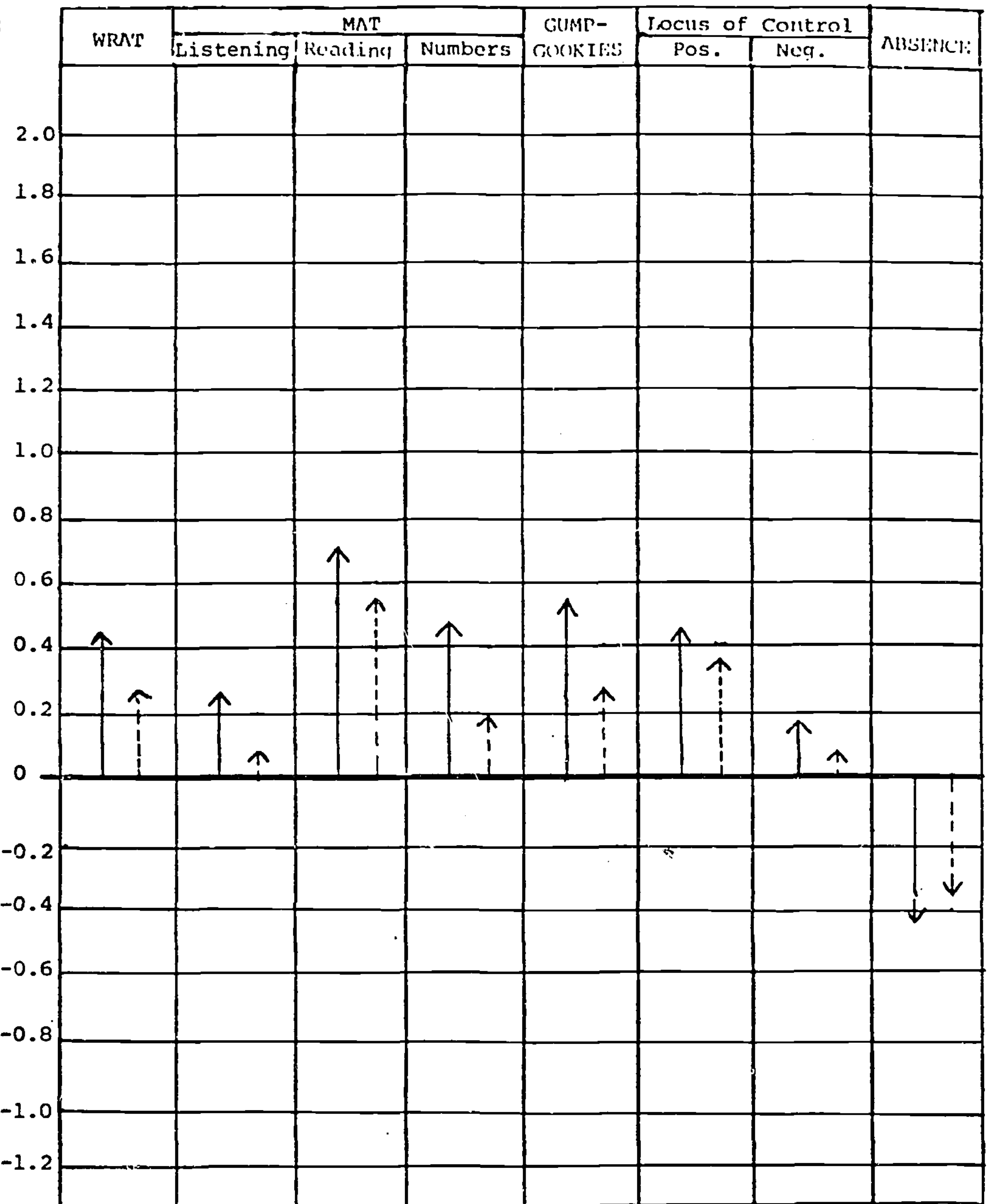
Figure VII-17 displays the school level FT/NFT contrasts for the High/Scope program on the Spring outcome measures. There is only one significant contrast: the FT group exceeds the NFT group on the MAT Reading subtest. However, there are trends favoring the FT group on all outcomes, except Locus of Control (negative).

Figure VII - 17
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 9

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)





		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.387	0.993	1.718	1.436	7.107	0.252	0.066	-2.385
	Unadj.	0.244	0.313	1.401	0.610	3.164	0.214	0.024	-2.090
SE = Standard Error of B	Adjusted	0.239	1.123	0.762	0.929	4.261	0.192	0.149	1.985
	Unadj.	0.329	1.341	0.919	1.118	4.864	0.232	0.147	2.266
t = B/SE = "Significance" Statistic	Adjusted	1.619	0.884	2.254	1.546	1.668	1.312	0.444	-1.201
	Unadj.	0.743	0.233	1.525	0.545	0.651	0.922	0.162	-0.922
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.438	0.267	0.706	0.463	0.567	0.444	0.188	-0.414
	Unadj.	0.276	0.084	0.576	0.196	0.253	0.377	0.068	-0.363
N = Number of Schools in Computation	FT	11	11	11	11	11	11	11	11
	NFT	12	12	12	12	12	12	12	12

4.6.2 Class Level FT/NFT Contrasts

The distribution of FT/NFT classes is similar to the distribution of schools for this Sponsor, except that there are no NFT classes in the large Northeastern city. Moreover, the demographic characteristics of the FT/NFT groups parallel those described at school level.

Figure VII- 31 displays the FT/NFT contrasts at the class level of analysis. The FT group exceeds the NFT group substantially on the Locus of Control measures and, to a lesser extent, on the WRAT and MAT Reading subtest as well. There are also very small positive trends on the other MAT subtests, but these trends are extremely small.

Thus, despite the geographic and demographic similarities in the groups analyzed at school and class levels, there are differences in the results found. With the exception of the Locus of Control measures, the school level contrasts more clearly favor the FT group. These differences are not easily interpretable with data currently available.

4.6.3 Child Level FT/NFT Contrasts

At the child level of analysis, the FT/NFT groups are distributed somewhat differently from the schools and classes. Well over half the FT/NFT schools and classes were drawn from the Western sites, but a smaller percentage of children were drawn from there.

There are also some shifts in the demographic characteristics of the FT/NFT groups at this level of analysis. As before, the FT group is lower in mean adjusted income and equal in mother's education to the NFT group. However, at this level of analysis the FT group exceeds the NFT group slightly on initial achievement. Furthermore, the disparity between the two groups in the percentage of minority children served is even larger (FT = 86%; NFT = 48%).

The child level contrasts are displayed in Figure VII- 32. As in the school level analyses, there is a positive trend in favor of the FT group on all achievement outcomes, absence, and Locus of Control (positive). The FT group also exceeds the NFT group on Locus of Control (negative),

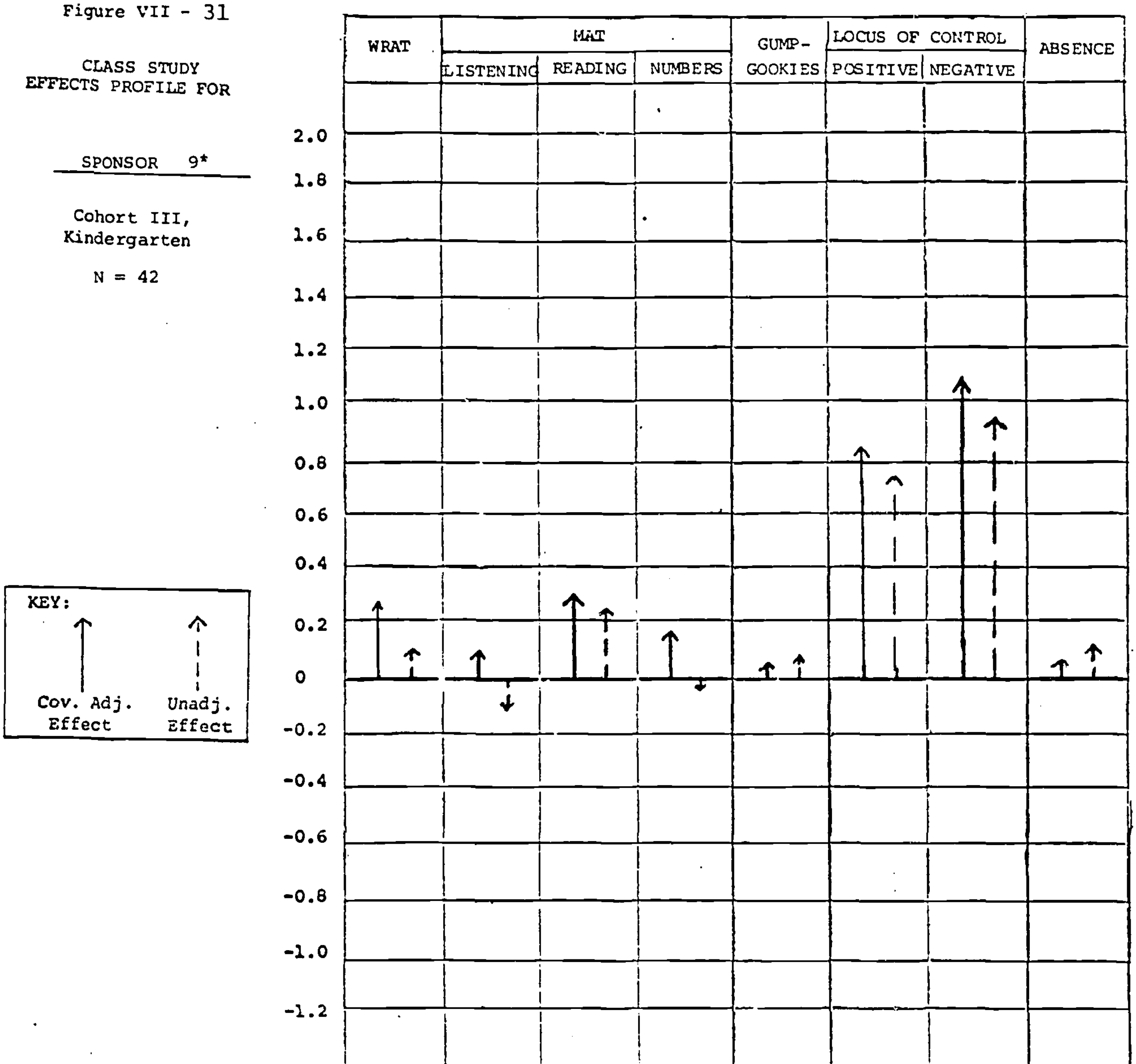
Figure VII - 31

CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 9*

Cohort III,
Kindergarten

N = 42



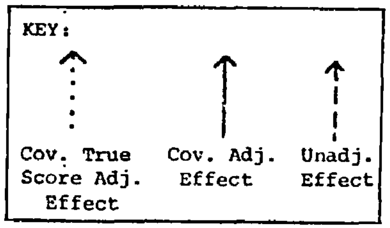
		WRAT	MAT Listening	MAT Reading	MAT Numbers	Gump-Gookies	Locus of Control Positive	Locus of Control Negative	Absence
B = Magnitude of Effect	Adjusted	1.636	.274	.891	.682	.054	.662	.567	.191
	Unadj.	.652	-.373	.691	-.007	.127	.598	.516	.620
*B/S = Effect in Standard Deviations	Adjusted	.229	.063	.294	.183	.016	.830	1.028	.032
	Unadj.	.091	-.085	.228	-.002	.037	.750	.936	.103
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

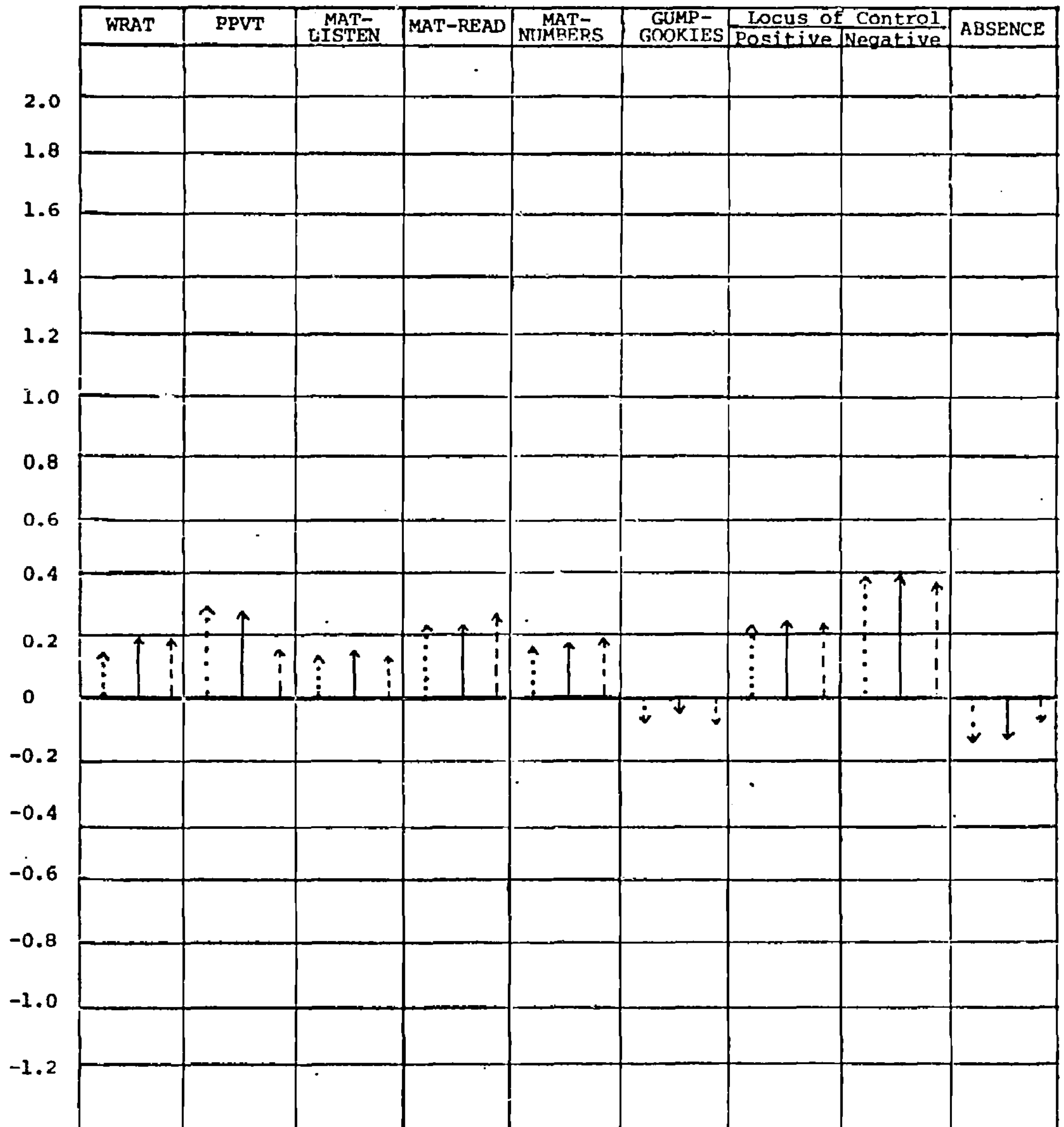
Figure VII - 32

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 9 *

Cohort III,
Kindergarten

N = 317





		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-GOOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	2.374	2.060	1.059	1.235	1.127	-.521	.469	.635	-1.723
	Adjusted	2.526	1.980	1.120	1.284	1.186	-.493	.478	.638	-1.736
	Unadjusted	2.512	1.120	1.046	1.371	1.214	-.636	.463	.606	-1.111
*B/S = Effect in Standard Deviations	True Score Adj.	.186	.270	.136	.220	.169	-.062	.232	.394	-.145
	Adjusted	.198	.260	.144	.229	.178	-.059	.237	.396	-.146
	Unadjusted	.197	.147	.135	.244	.182	-.076	.230	.376	-.094
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

which parallels the class finding, and on the PPVT. There is no difference between the two groups on the Gumpgookies test.

4.6.4 Selected Teacher Data

The FT teachers for High/Scope are average in age, education, and overall teaching experience, compared to the total group of FT teachers for all Sponsors. They receive higher salaries, however, than any other Sponsor group. Although they have taught for an average number of years overall, they are relatively new to their current school assignments and have been with the Sponsor for a shorter period of time than any other Sponsor group. Approximately 36% of these FT teachers are from minority groups.

The NFT teachers for this Sponsor are a great deal older and more experienced than the FT teachers. In fact, the High/Scope NFT teachers are one of the most experienced and stable NFT teacher groups. The NFT group does not differ greatly from the FT group, however, in educational attainment, in salary, or in ethnicity.

High/Scope teachers report receiving relatively little training overall. What training does occur is primarily in child-centered learning activities.

The FT teachers in the High/Scope program are the most child-centered teachers in the sample, compared to their NFT counterparts as well as to other Sponsor groups. The FT teachers do not differ from the NFT teachers in the values they place on parent involvement; however, the NFT group is higher than any other on this variable. Moreover, the FT group makes more home visits than the NFT group.

Perhaps because of their relative inexperience, the High/Scope teachers are somewhat less satisfied with their Sponsor than the average FT teacher.

4.6.5 Summary and Discussion

The High/Scope program appears to be having some success in the

development of achievement, motivation, internal locus of control, and verbal ability as measured by this test battery. It also appears to be having some impact on attendance. Except for the motivation measure, these results are consistent across at least two levels of analysis in which there are some shifts in the geographic and demographic characteristics of the samples.

It may be, however, that something in the composition of the classes analyzed differentially affects the way in which this program works, for it is at the class level that the results are most inconsistent. In this report and in future studies, we will explore classroom/teacher characteristics in greater depth to determine the classroom contexts in which this program has most success.

4.7.0 SPONSOR 10: UNIVERSITY OF FLORIDA PARENT EDUCATION MODEL

The University of Florida program is described as a Parent Education Model. Based on the premise that children's learning takes place as much at home as in school, the major objective is to improve children's school achievement through educating parents to participate directly in the education of their children. While the curriculum is not standardized, it does have a Piagetian orientation.

4.7.1 School Level FT/NFT Contrasts

The subset of University of Florida schools was drawn from five sites located in all four geographic areas. Over half of the FT schools and three-fourths of the NFT schools are located in one large Southern site.

Compared to the total group of FT schools for all Sponsors, the University of Florida schools are average in the adjusted income level of the families served. On the other hand, they are below average, relative to the total FT group, on mother's education level and initial achievement, as measured by the Fall WRAT.

Compared to their NFT group, the University of Florida FT schools are slightly lower in mean adjusted income level. The two groups are similar, however, in the mean percentage of mother's completing high school and in the mean percentage of minority pupils served (FT = 65%, NFT = 60%). Furthermore, the FT group is higher than the NFT group on entering achievement scores. Overall, therefore, the FT group has a slight initial advantage over the NFT group.

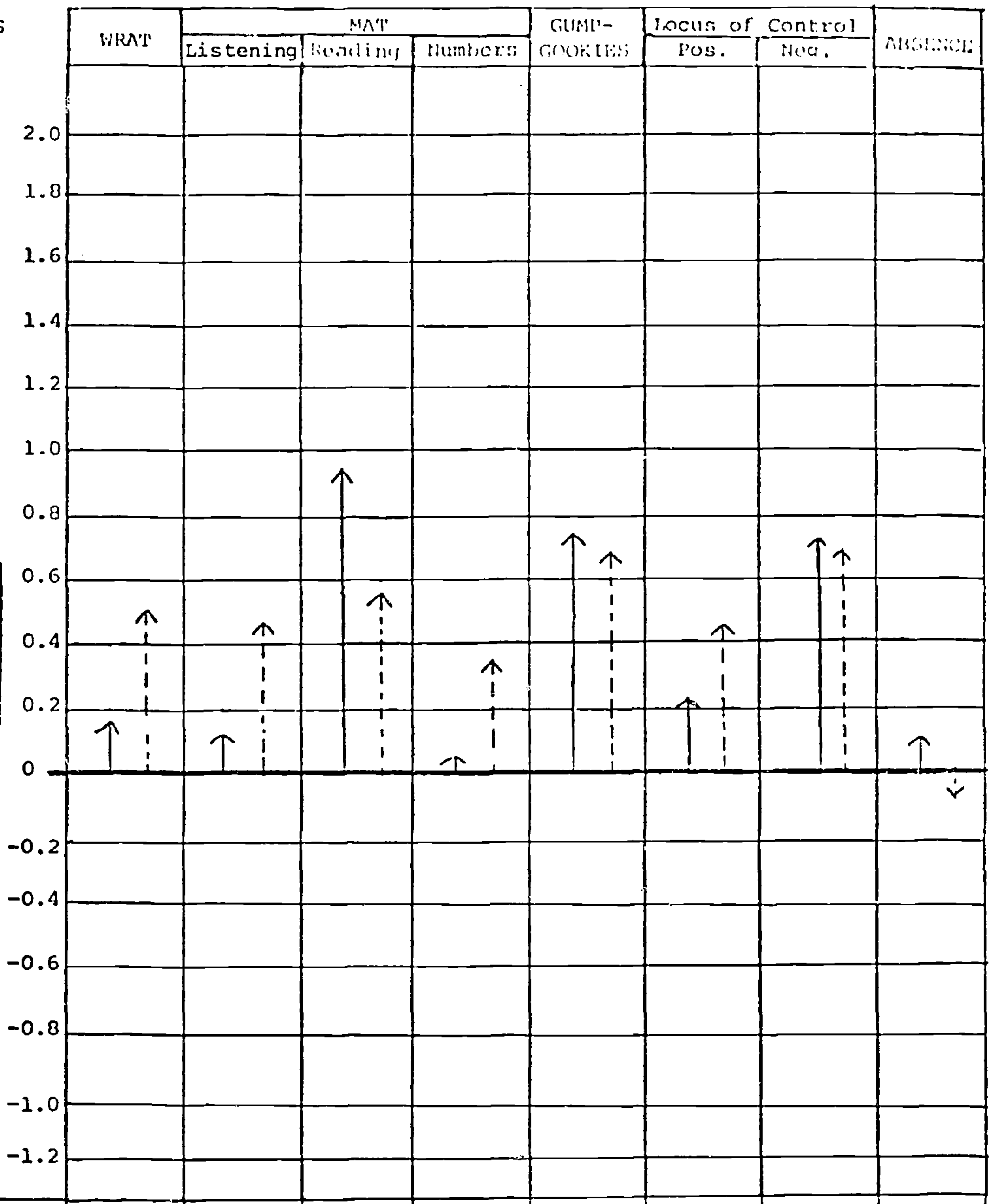
The school level FT/NFT contrasts are displayed in Figure VII-18. With initial differences partialled out, there is only one significant contrast: the FT schools exceed the NFT schools on the Gumpgookies test. In addition there are trends which favor the FT group on the MAT reading subtest and on the Locus of Control measures.

Figure VII - 18
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 10

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:
Cov. Adj. Effect
Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.150	0.551	0.711	0.155	9.569	0.132	0.262	0.695
	Unadj.	0.463	1.743	1.421	1.139	8.498	0.253	0.252	-0.324
SE = Standard Error of B	Adjusted	0.237	1.112	0.755	0.920	4.220	0.190	0.147	1.966
	Unadj.	0.325	1.324	0.907	1.032	4.803	0.229	0.145	2.238
t = B/SE = "Significance" Statistic	Adjusted	0.636	0.495	0.942	0.168	2.268	0.693	1.782	0.353
	Unadj.	1.426	1.316	1.566	1.103	1.769	1.106	1.735	-0.145
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	0.170	0.148	0.942	0.050	0.763	0.232	0.745	0.121
	Unadj.	0.524	0.468	0.584	0.367	0.678	0.446	0.716	-0.056
N = Number of Schools in Computation	FT	17	17	17	17	17	17	17	17
	NFT	9	9	9	9	9	9	9	9

4.7.2 Class Level FT/NFT Contrasts

The subset of classes included in these analyses is similar in geographic distribution to the subset of schools. Over half the FT classes and two-thirds of the NFT classes are located in the large Southern city.

The SES differences between the FT/NFT groups are greater at the class than at the school level. The NFT group is not only higher in adjusted income level, but also in mother's education. Furthermore, there is a greater disparity between the two groups in the mean percentage of minority pupils (FT = 64%, NFT = 47%). On the other hand, the FT group exceeds the NFT group on entering achievement level, which parallels the school level.

Figure VII-33 displays the class level FT/NFT contrasts. The FT classes exceed the NFT classes on all achievement outcomes as well as on the Gumpgookies test and Locus of Control (negative). There is also a slight trend in favor of the FT group on Locus of Control (positive), but this is not significant. The school and class results are similar in some respects, dissimilar in others. The FT group exceeds the NFT group more consistently in the achievement domain at the class level, and less consistently in the affective area. These differences may reflect differences in the demographic characteristics of the two groups at the school and class levels.

4.7.3 Child Level FT/NFT Contrasts

The distribution of FT/NFT children at the child level of analysis differs from the distribution of schools and classes. Only 37% of the FT children and 48% of the NFT children are located in the Southern site. A larger percentage of FT and NFT children are located in the North Central site.

The FT group exceeds the SES group on both SES and percentage of minority pupils, differences which parallel those found at the class level. On the other hand, the FT group is slightly lower than the NFT group on entering achievement, while the reverse is true at both class and school levels.

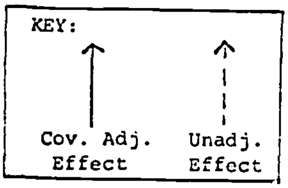
Figure VII - 33

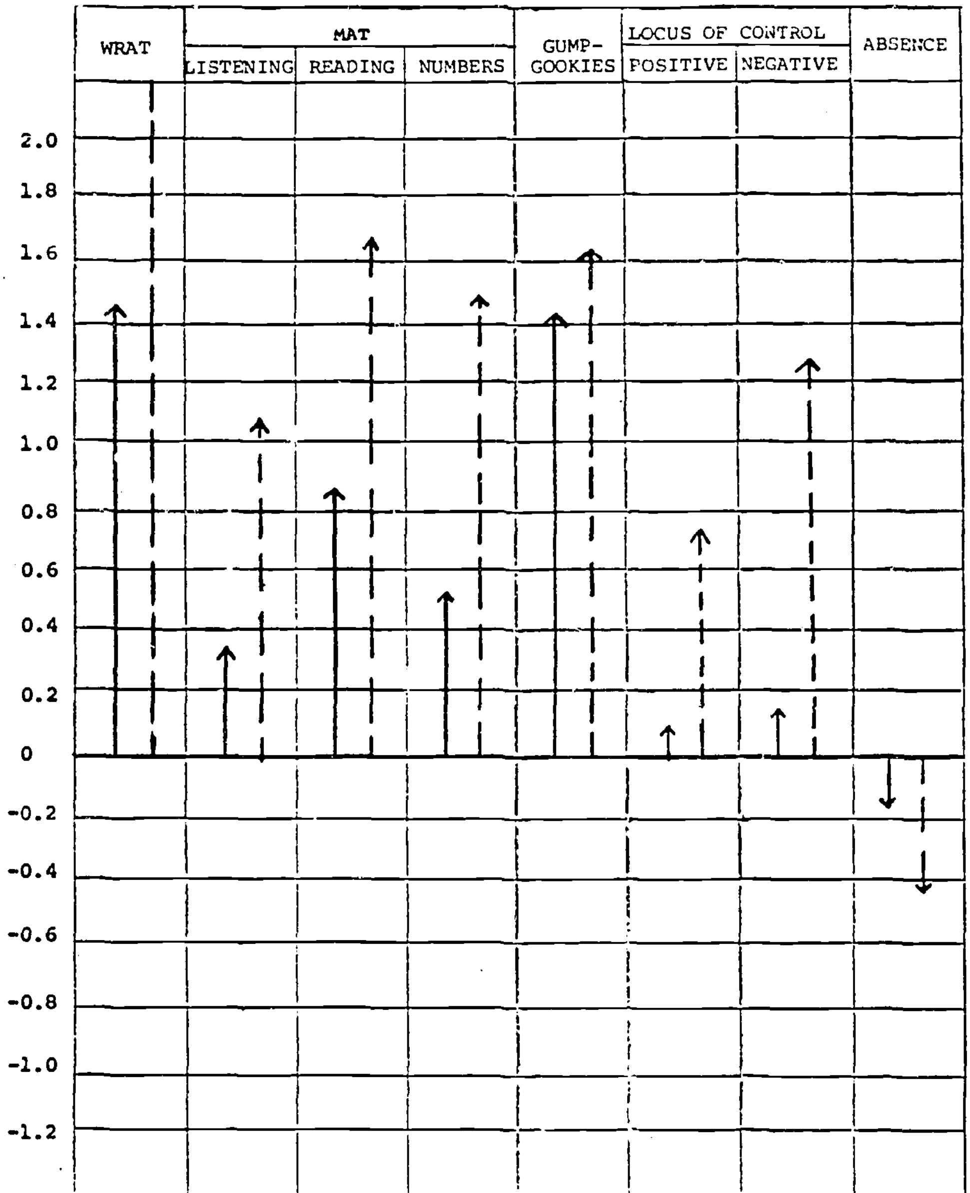
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 10*

Cohort III,
Kindergarten

N = 32





		WRAT	LISTENING	READING	NUMBERS	GUMP-GOOKIES	POSITIVE	NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	10.397	1.520	2.510	2.017	4.781	.065	.748	-1.082
	Unadj.	18.296	4.874	4.949	5.481	5.498	.606	.716	-2.505
* B/S = Effect in Standard Deviations	Adjusted	1.454	.348	.829	.539	1.406	.081	1.357	-.179
	Unadj.	2.558	1.115	1.634	1.466	1.617	.760	1.299	-.414
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

Figure VII-34 displays the FT/NFT contrasts for the University of Florida program at the child level of analysis. At this level, the FT group exceeds the NFT group on all achievement outcomes, the Gumpgookies test, and the Locus of Control (negative) measure. These results parallel those found at class level, and are stronger than those found at school level. In addition, the FT group exceeds the NFT group significantly on the PPVT.

4.7.4 Selected Teacher Data

The FT teachers in the University of Florida program are slightly above average in age and overall teaching experience, compared to other FT teachers. They are also similar in education and salary level. A relatively high proportion (42%) of these teachers are from minority groups.

Unlike most other FT/NFT teacher groups, the NFT teachers for this Sponsor are younger and less experienced than the FT teachers. The two groups are similar in education, salary, and ethnicity.

The Florida teachers report receiving far less training in structured or child-centered learning activites than other Sponsor groups. On the other hand, they report receiving far more training in working with parents and aides.

These FT teachers place slightly more value on involving parents than their NFT counterparts, and far more than other Sponsor FT groups. They also make more home visits compared to both their NFT counterparts and other FT teachers.

4.7.5 Summary and Discussion

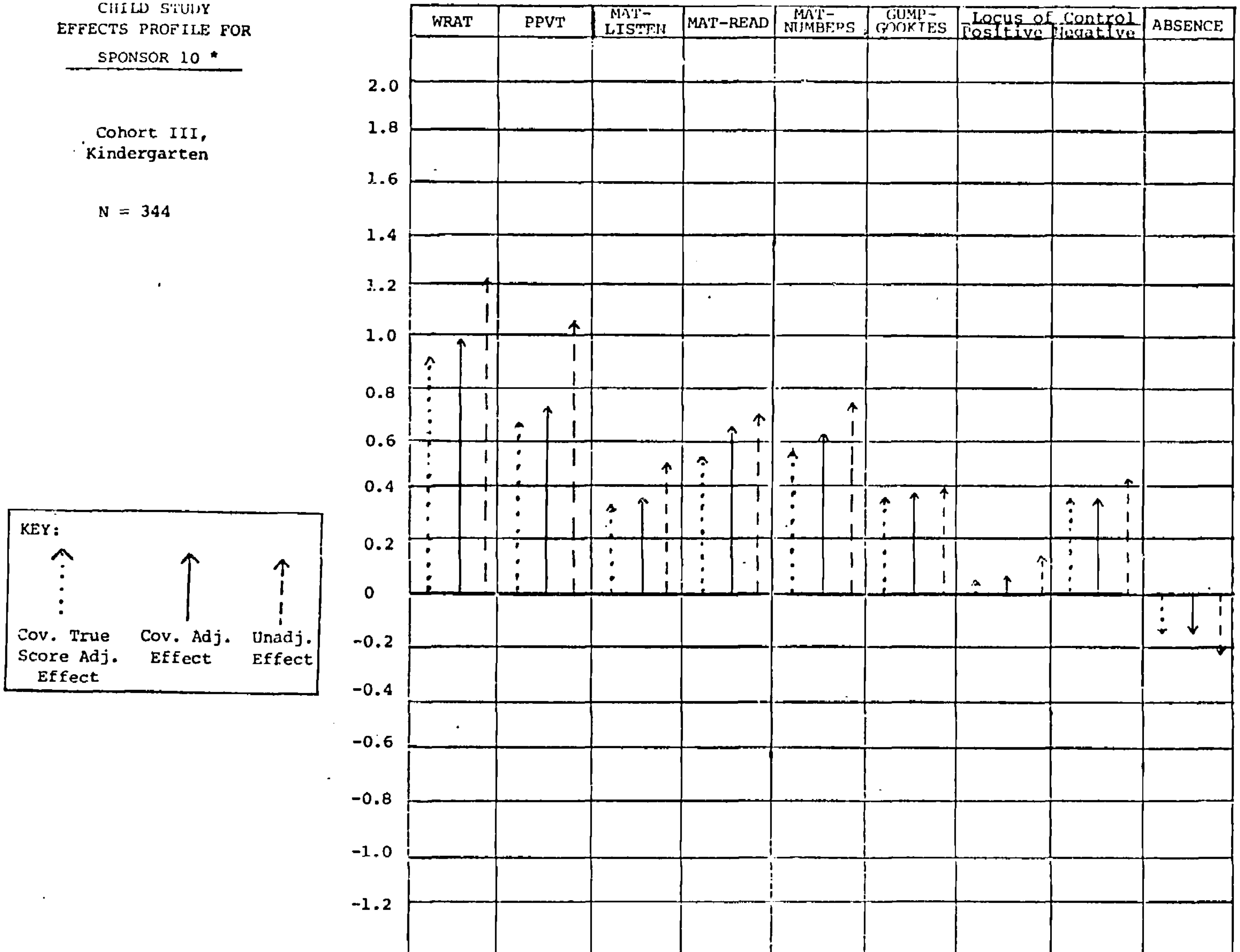
The University of Florida program appears to be having positive effects on reading achievement at the school level analyses, and on a variety of achievement outcomes at class and child levels. The FT group also exceeds the NFT group at all three levels of analyses on achievement motivation, as measured by the Gumpgookies. Given that there are differences in the characteristics of the children and the communities included at the different levels, these findings suggest

Figure VII - 34

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 10 *

Cohort III,
Kindergarten

N = 344



		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-COOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	11.933	5.218	2.673	2.958	3.862	3.023	.109	.602	-1.822
	Adjusted	12.635	5.855	2.954	3.684	4.131	3.154	.152	.618	-1.884
	Unadjusted	15.420	8.260	3.983	4.059	5.166	3.341	.302	.659	-2.436
*B/S = Effect in Standard Deviations	True Score Adj.	.936	.685	.344	.526	.579	.361	.054	.374	-.154
	Adjusted	.991	.769	.380	.656	.619	.376	.076	.374	-.159
	Unadjusted	1.209	1.085	.513	.722	.774	.399	.150	.409	-.205
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

that the Florida program may be robust in producing achievement and motivation effects across a variety of settings.

In addition, the Florida program also appears to be having an impact on the PPVT which has been found to be correlated with other measures of intelligence.

4.8.0 SPONSOR 11: THE EDUCATIONAL DEVELOPMENT CENTER (EDC)

The EDC program is based on the belief that learning is facilitated by a child's active participation in the learning process. In the flexible, open classroom environment, children are encouraged to initiate activities, pursue their interests, and generally assume responsibility for their own learning. The basic objective of the program is to provide the optimal environment for children's growth in academic and problem solving skills, self expression and self direction.

4.8.1 School Level FT/NFT Contrasts

The subset of EDC schools included in these analyses are located in four Northeastern sites, with the exception of a single FT school located in a large Southern city. Over 60% of the FT schools and 50% of the NFT schools are in two medium sized Northeastern cities.

Compared to the total group of schools for all Sponsors, these EDC schools serve relatively high SES families. These FT schools are higher than those of any other Sponsor group in adjusted income level and in mean percentage of mothers completing high school. The schools are also higher than average in entering achievement. Finally, the mean percentage of minority pupils is relatively low, compared to other groups.

Although the FT schools for this Sponsor are relatively high on SES and entering achievement compared to other Sponsor groups, they are not as high on these indices as the NFT schools. The NFT schools are higher on adjusted income and entering achievement. They are also somewhat higher on mother's education, although this difference is slight. The mean percentage of minority pupils in FT schools (45%) is somewhat higher than for NFT schools (36%).

Figure VII-19 displays the school level FT/NFT contrasts for the EDC program on each of the Spring outcome measures. With initial differences taken into account, the FT group exceeds the NFT group on the Gumpgookies

Figure VII - 19

FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 11

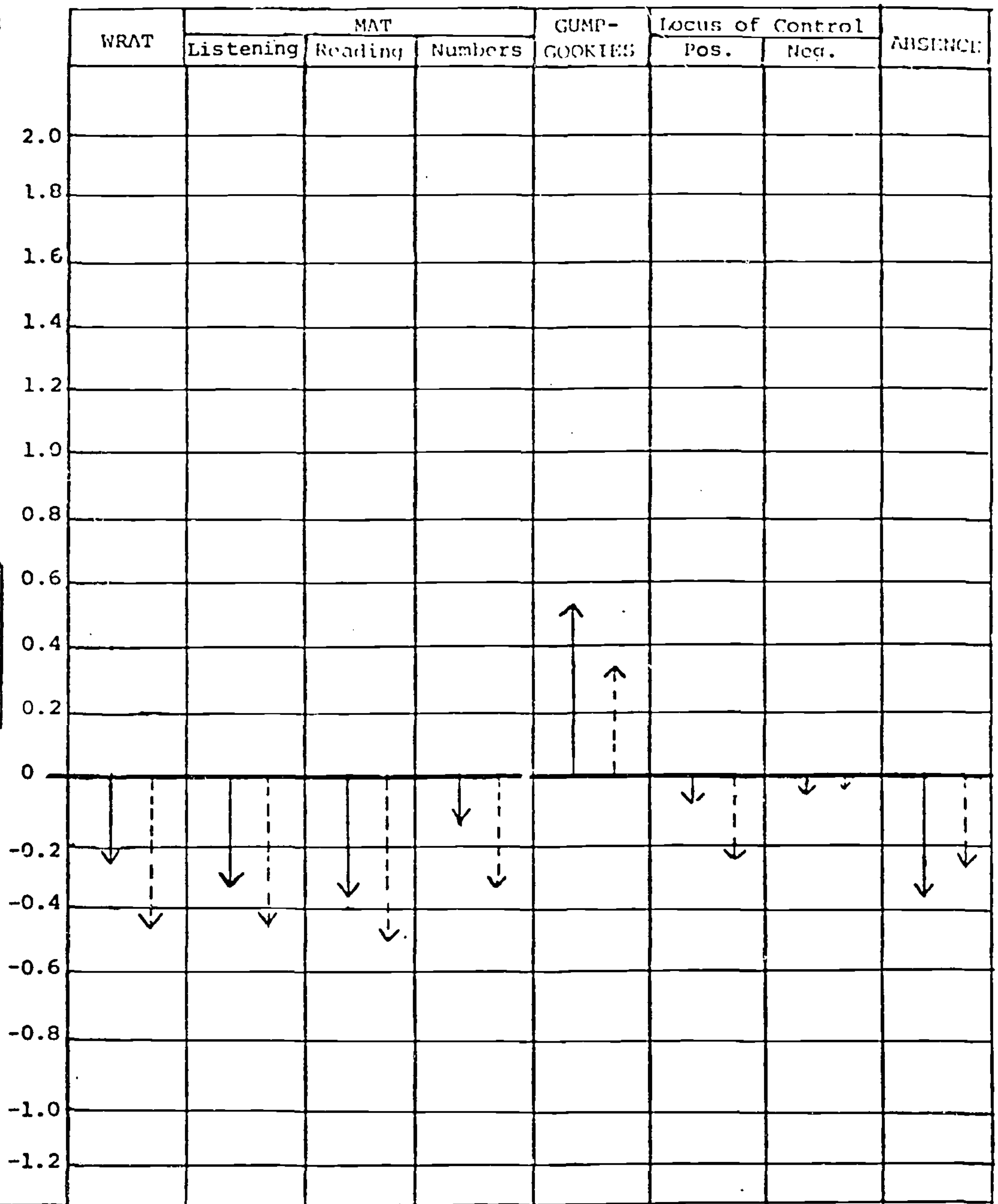
Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect

Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.205	-1.274	-0.901	-0.435	6.775	-0.042	-0.014	-2.162
	Unadj.	-0.382	-1.645	-1.153	-0.993	4.469	-0.143	-0.009	-1.482
SE = Standard Error of B	Adjusted	0.225	1.057	0.717	0.874	4.008	0.181	0.140	1.868
	Unadj.	0.315	1.286	0.881	1.072	4.665	0.222	0.141	2.174
t = B/SE = "Significance" Statistic	Adjusted	-0.914	-1.206	-1.257	-0.498	1.690	-0.230	-0.099	-1.158
	Unadj.	-1.211	-1.279	-1.308	-0.926	0.958	-0.643	-0.063	-0.682
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S = Effect in Standard Deviations	Adjusted	-0.232	-0.342	-0.370	-0.140	0.541	-0.073	-0.039	-0.375
	Unadj.	-0.432	-0.442	-0.474	-0.320	0.357	-0.252	-0.025	-0.257
N = Number of Schools in Computation	FT	13	13	13	13	13	13	13	13
	NFT	12	12	12	12	12	12	12	12

test and the measure of attendance. On the other hand, the NFT group scores higher on the WRAT and MAT subtests. However, none of these differences are statistically significant.

4.8.2 Class Level FT/NFT Contrasts

The EDC classes were drawn from the same sites as the schools, with the exception that there are also NFT schools in the large Southern site. However, the distribution of classes by site is very different. Only 35% of the FT group and 40% of the NFT group are located in the middle-sized Northeastern site, far less than at school level. On the other hand, there is a greater proportion of classes in each of the other sites at this level.

The demographic differences between the FT/NFT classes are similar to those for the FT/NFT schools. The NFT classes are higher than the FT classes on adjusted income level and entering achievement. The two groups are similar in mother's education. The FT group has a higher mean percentage of minority children (64%) than the NFT group (50%).

Figure VII-35 displays the FT/NFT contrasts for the class level analyses. At the class level, the FT group exceeds the NFT group on the Gumpgookies and the NFT group exceeds the FT group on the WRAT. These results parallel those found at the school level. On the other hand, the directions of the absence and MAT Listening contrasts are reversed, the former favoring the NFT group and the latte' the FT group.

The most marked difference between the two levels is the substantial shift in the absence measure from school to class level. Since the Gumpgookies contrasts do not shift from one level to another, it does not appear that children's motivation is differentially affecting the absence rate. However, there may be other health, climate, or parental factors which differ by communities and are reflected in these attendance measures.

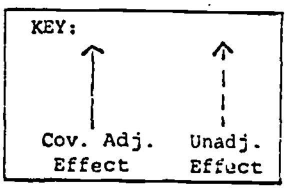
Figure VII - 35

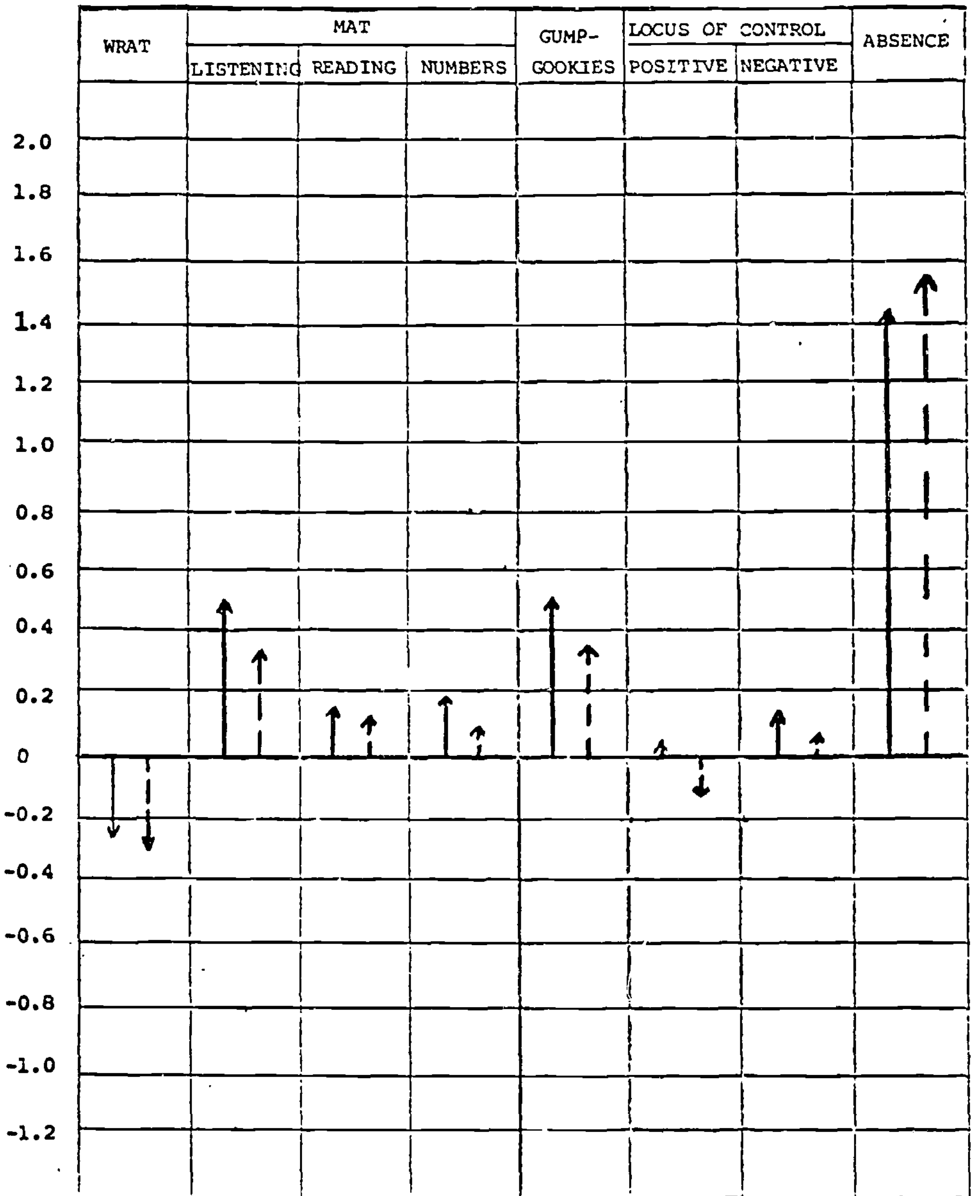
CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 11*

Cohort III,
Kindergarten

N = 35





		WRAT	LISTENING	READING	NUMBERS	GUMP-GOOKIES	POSITIVE	NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	-1.649	2.203	.493	.709	1.680	.015	.077	8.480
	Unadj.	-2.071	1.368	.377	.265	1.248	-.088	.022	9.552
*B/S = Effect in Standard Deviations	Adjusted	-.231	.504	.163	.190	.494	.019	.139	1.402
	Unadj.	-.290	.313	.125	.071	.367	-.111	.040	1.579
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

4.8.3 Child Level FT/NFT Contrasts

The distribution of FT children at this level of analysis is similar to the distribution of FT classes, but the distribution of NFT children is markedly different.

Compared to the NFT class distribution, very few NFT children are located in the large Northeastern site, and none in the Southern site. The net effect of these shifts is to 1) make cross level comparisons inappropriate, and 2) introduce a geographic mismatch within the child level sample.

There are also shifts in the demographic differences between the two groups. The FT children are equal in adjusted income level to the NFT children, but lower in mother's education, whereas at school and class levels the reverse is true. Moreover, there is a greater disparity in percent minority children at this level; over 48% of the FT group and only 23% of the NFT group are minority children. On the other hand, the NFT children exceed the FT children on entering achievement levels, which parallels the other groups.

Figure VII-36 displays the child level FT/NFT contrasts for EDC. When initial differences are statistically adjusted the FT group has a small advantage on the MAT Reading subtest. The NFT group has a small advantage on the absence measure and the WRAT, which parallels the results of the class analysis. However, there is no longer an FT favoring trend on the Gumpgookies.

4.8.4 Selected Teacher Data

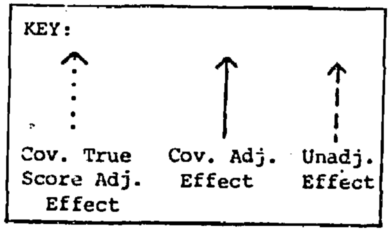
The EDC FT teachers are average, or slightly above, in age, experience and salary relative to other FT teachers. Three quarters of the EDC teachers have obtained advanced credits or degrees, which also parallels the overall FT group. The FT teachers for this Sponsor are predominantly White.

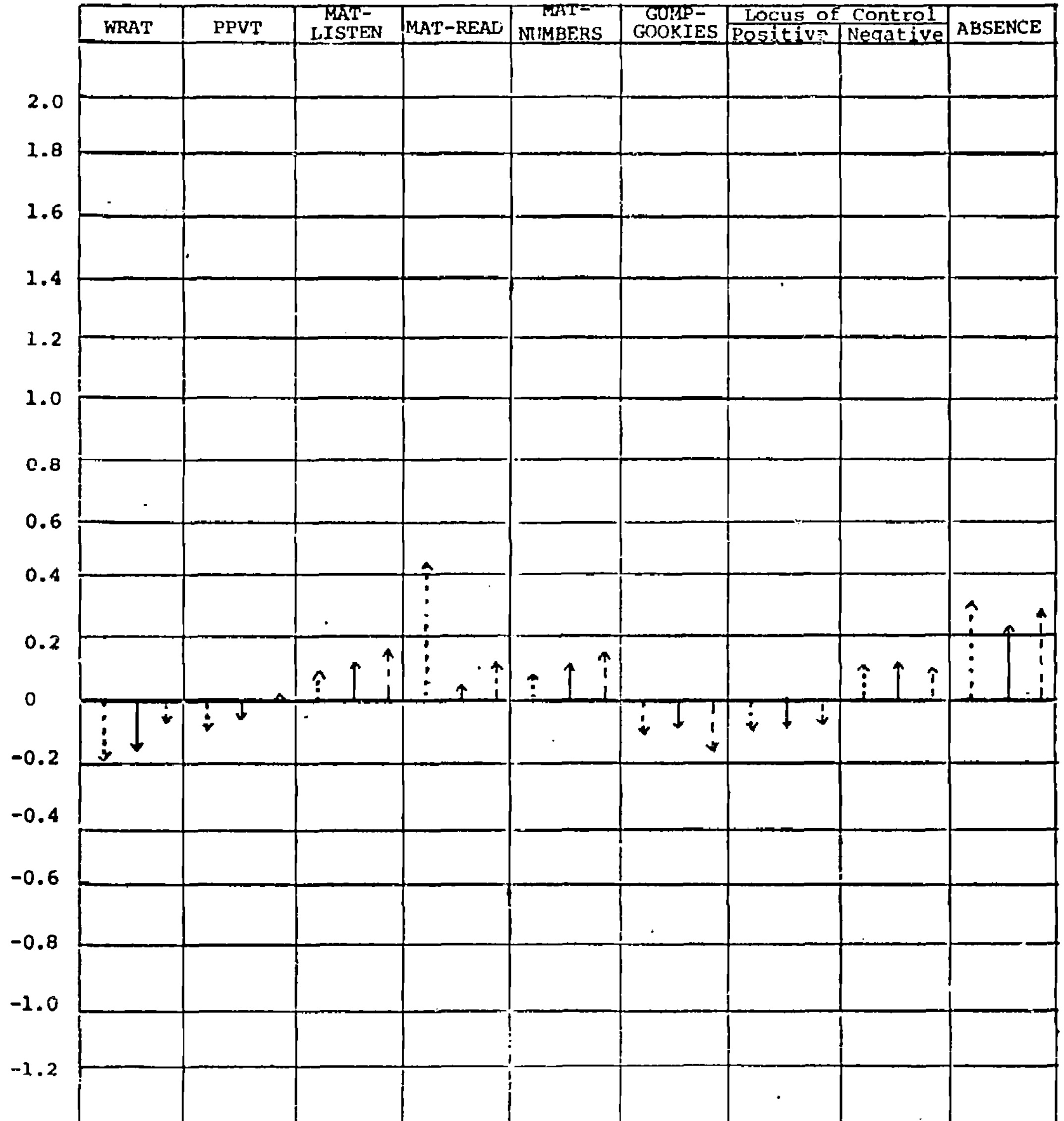
The NFT teachers are older, more experienced, and more highly educated than the FT teachers for this Sponsor. They also receive higher

Figure VII - 36

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 11*

Cohort III,
Kindergarten
N = 397





		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-GOOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	-2.421	-.797	.784	2.400	.577	-1.003	-.240	.207	3.675
	Adjusted	-2.037	-.555	.938	.219	.724	-.931	-.216	.216	2.588
	Unadjusted	-1.097	.223	1.089	.709	.865	-1.259	-.182	.183	3.929
*B/S = Effect in Standard Deviations	True Score Adj.	-.190	-.103	.101	.427	.086	-.120	-.119	.129	.310
	Adjusted	-.160	-.073	.121	.039	.108	-.111	-.107	.134	.218
	Unadjusted	-.086	.029	.140	.126	.130	-.150	-.090	.114	.331
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

salaries than the EDC FT group. There are far more minority teachers in the NFT group (47%) than in the FT group (6%).

FT teachers in the EDC program report receiving a great deal of Sponsor training in child-centeredness, but little in other areas. In turn, they place more value on children's exploring and manipulating of their environment and less on social skills development relative to their NFT counterparts. They are not very different from the NFT group in either their values or behaviors toward parents.

Finally, the EDC FT teachers are relatively more satisfied with their Sponsor's approach than the average FT teacher.

4.8.5 Summary and Discussion

While it appears that the EDC program is having some impact on certain children in both achievement and motivation, it varies greatly depending upon the analytic sample.

An exploration of the characteristics of the samples for these three studies suggests that geographical factors may account for these discrepancies. The three samples were all drawn from the same sites, but each differed in the relative proportion of the FT/NFT groups located in these sites. The differences in geographic distribution were not matched by sharp differences in the characteristics of the samples, except for the percentage of minority children included. However, they may reflect differences in community characteristics or in program implementation not yet examined.

The EDC program, being concerned with the process of learning as much as if not more than the product, is perhaps more susceptible to differences in implementation than any other. We have found that the EDC teachers, in general, value the program's goals. In future studies we will explore whether variation in the implementation of those goals affects pupil performance.

4.9.0 SPONSOR 12: UNIVERSITY OF PITTSBURGH, PRIMARY EDUCATION PROJECT

The Primary Education Project utilizes a number of interrelated curriculum components which are carefully structured and sequenced to provide for optimally efficient learning. Three general classes of skills are included in these curriculum components which are designed to form the foundation of all higher level functioning: (1) orienting and attending skills, (2) perceptual motor skills, and (3) conceptual-linguistic skills. These latter include classification, reasoning, memory, language, and mathematics concepts. The curriculum is highly individualized in order to allow the child to progress at his own pace. The teacher serves as a facilitator and resource person as the child moves through each component.

4.9.1 School Level FT/NFT Contrasts

The subset of schools included in the school level analysis for this Sponsor was drawn from three sites, two in North Central United States and one in the Northeast. The NFT schools are distributed evenly among these sites. The FT schools are distributed less evenly. Approximately 22% of the FT schools are located in the large North Central site, 44% in the small Northeastern site, and 33% in the rural, North Central site.

First, we shall compare the subset of this Sponsor's FT schools to the total group of FT schools for all Sponsors. Although there is a great deal of variability among the three sites, the FT schools for this Sponsor, on the average, are similar in adjusted income level to the schools for all other Sponsors. However, the mean percentage of mothers completing high school is higher for the University of Pittsburgh schools than for any other Sponsor group. In addition, the FT group has a higher mean percentage of White pupils and a higher mean score on the Fall WRAT than any other Sponsor group. Thus, overall, the University of Pittsburgh schools in this sample serve children and families relatively high on the scale of demographic characteristics, when compared with other schools.

Next we shall compare the FT/NFT schools for this Sponsor. On the average, the FT schools in the Pittsburgh program serve families of lower adjusted income than the NFT schools. However, the two groups are relatively well matched on the other demographic indices. The mothers of the children in the FT schools have achieved, on the average, the same educational level as the NFT school mothers. Also, the mean percentage of White pupils for the FT schools (78%) is similar to that for the NFT schools (72%).

Finally, the two groups enter school with approximately the same achievement levels, as measured by the Fall WRAT.

Figure VII-10 presents the school level FT/NFT contrasts for the Pittsburgh program on the Spring outcome measures. With initial differences partialled out, there are three significant FT/NFT contrasts: the FT group exceeds the NFT group on the MAT arithmetic subtest, on the Gumpgookies test, and the Locus of Control (negative) measure. There are also trends in favor of the FT group on the WRAT and the Locus of Control (positive) measure. There are no important differences between the two groups on the other measures at this level of analysis.

4.9.2 Class Level FT/NFT Contrasts

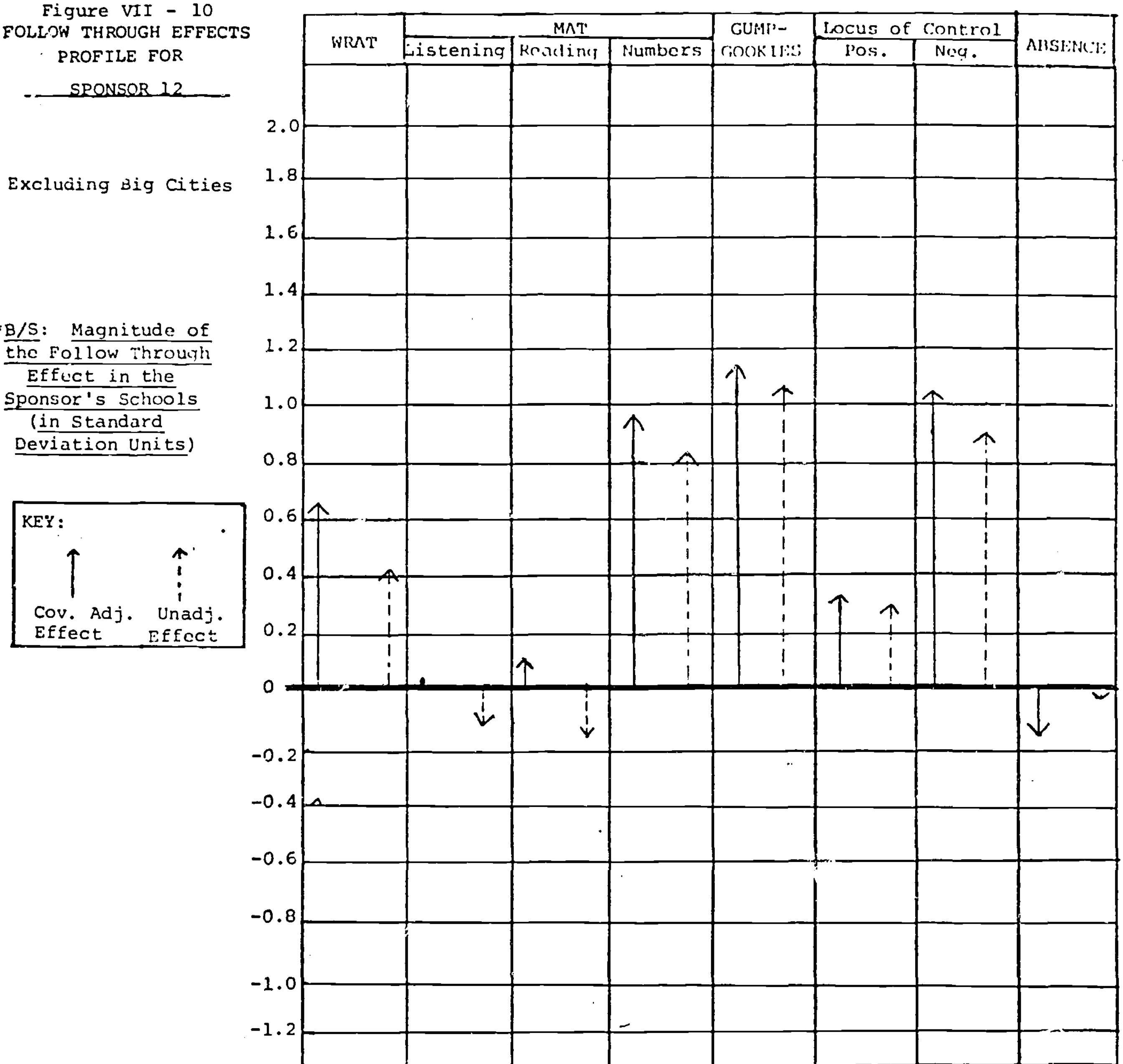
The group of classes which were included in these analyses was drawn from the same sites as the group of schools. However, the distribution of classes by site is somewhat different from the distribution of schools. For example, a larger percentage of both FT classes (40%) and NFT classes (46%) are in the rural, North Central site than the percentage of schools. (FT=33%; NFT=27%).

There are also certain demographic differences between the FT/NFT groups at the class level of analysis which do not parallel those at the school level of analysis. At both levels of analysis, the FT group is lower than the NFT group on adjusted income level. However, although the two groups are similar in the mean percentage of mothers completing high school at the class level, the FT group is lower than the NFT group on this measure. In addition, although the two groups are still approximately equal in ethnic composition at the class level, the relative percentage of White pupils in FT/NFT classes is slightly

Figure VII - 10
FOLLOW THROUGH EFFECTS
PROFILE FOR
SPONSOR 12

Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)



		WRAT	MAT Listening	MAT Reading	MAT Numbers	GUMP-GOOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	0.562	0.003	0.274	3.059	12.991	0.172	0.341	-0.799
	Unadj.	0.359	-0.396	-0.346	2.611	12.039	0.170	0.302	-0.101
SE = Standard Error of B	Adjusted	0.253	1.187	0.824	1.038	3.907	0.202	0.153	1.819
	Unadj.	0.354	1.447	1.004	1.243	4.242	0.243	0.149	2.108
t = B/SE = "Significance" Statistic	Adjusted	2.222	0.003	0.332	2.946	3.325	0.852	2.229	-0.439
	Unadj.	1.015	-0.274	-0.345	2.101	2.838	0.700	2.027	-0.048
S = Standard Deviation		0.872	3.742	2.434	3.157	11.279	0.549	0.330	5.014
*B/S = Effect in Standard Deviations	Adjusted	0.645	0.0008	0.113	0.969	1.152	0.314	1.033	-0.159
	Unadj.	0.412	-0.106	-0.142	0.827	1.067	0.309	0.913	-0.020
N = Number of Schools in Computation	FT	9	9	9	9	9	9	9	9
	NFT	11	11	11	11	11	11	11	11

different. At the class level the mean percentage of White children in FT classes is 70% and in NFT classes it is 77%. Finally, at the school level there was no difference between the two groups in entering achievement levels; however, at the class level, the FT group exceeds the NFT group slightly on Fall WRAT scores. Thus, at the class level of analysis, the FT group appears to be somewhat lower than the NFT group on SES measures and somewhat higher in initial achievement.

Figure VII-37 presents the class level FT/NFT contrasts for this Sponsor. At this level, the contrasts favor the FT group on all but one of the outcome variables. The FT group exceeds the NFT group on all achievement and affective tests except for the Gumpgookies test, where there is no significant difference between the two groups. The children in FT classes for this Sponsor are also absent 2.1 fewer days on the average than the children in NFT classes.

4.9.3 Child Level FT/NFT Contrasts

At the child level of analysis, approximately 44% of the FT children and 57% of the NFT children were drawn from the rural North Central site. The remaining children were drawn from the other two sites.

While there are some shifts in the magnitude of the differences between the FT/NFT groups, in general the child level FT/NFT sample is similar to the class level FT/NFT sample described above. The FT group is lower than the NFT group in SES, as measured by adjusted income and mother's education, but higher in initial achievement, as measured by the Fall WRAT. In addition, while both groups are predominantly White, the NFT group has a somewhat lower percentage of White children (73%) than the FT group (86%).

Figure VII-38 presents the FT/NFT contrasts for the child level of analysis. Although the contrasts are somewhat smaller than those found at class level, the direction of the contrasts is the same. All contrasts favor the FT group, except for the Gumpgookies test, where there are no differences between the two groups.

4.9.4 Selected Teacher Data

The FT teachers in the Pittsburgh program are slightly older

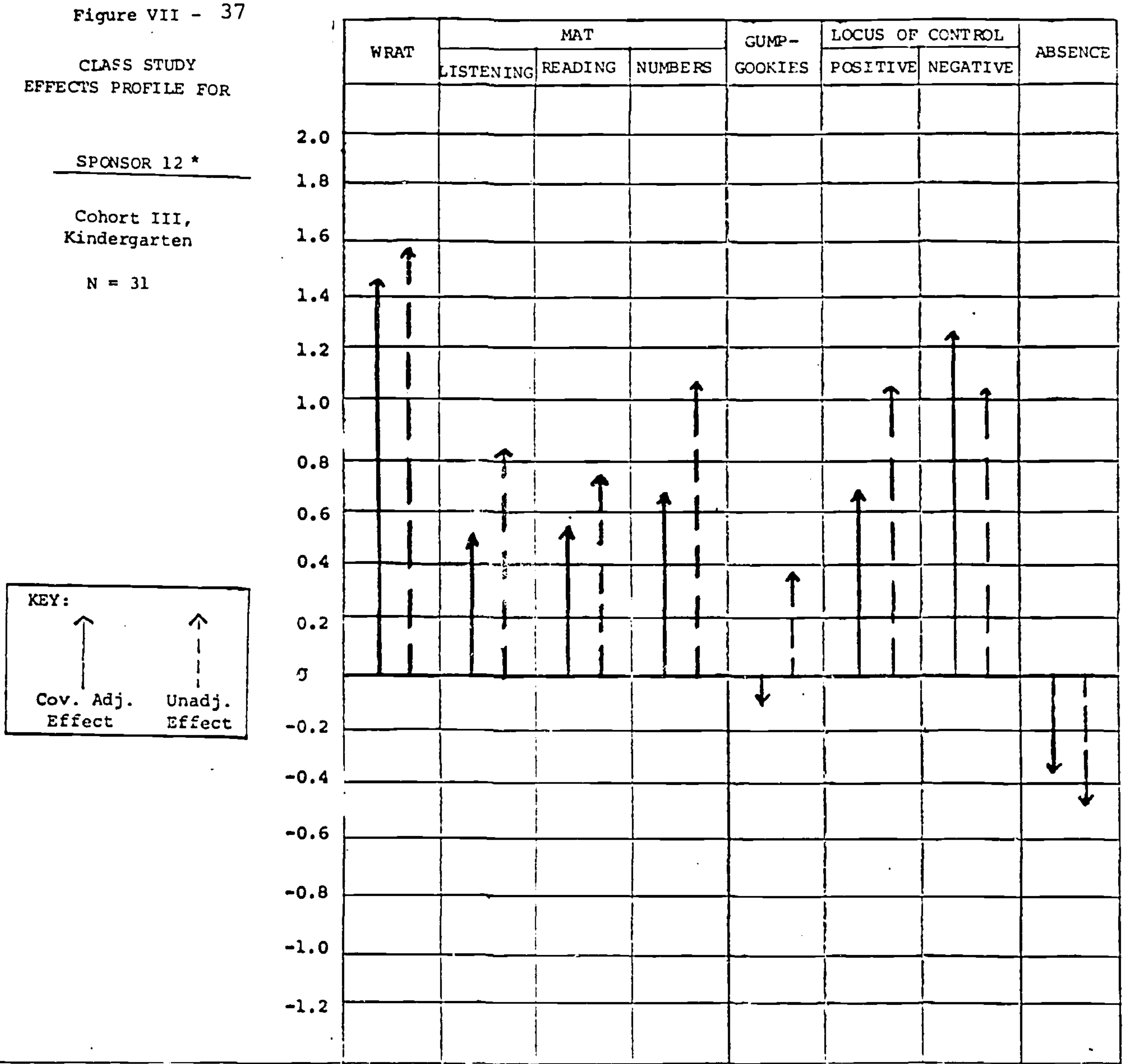
Figure VII - 37

CLASS STUDY
EFFECTS PROFILE FOR

SPONSOR 12 *

Cohort III,
Kindergarten

N = 31



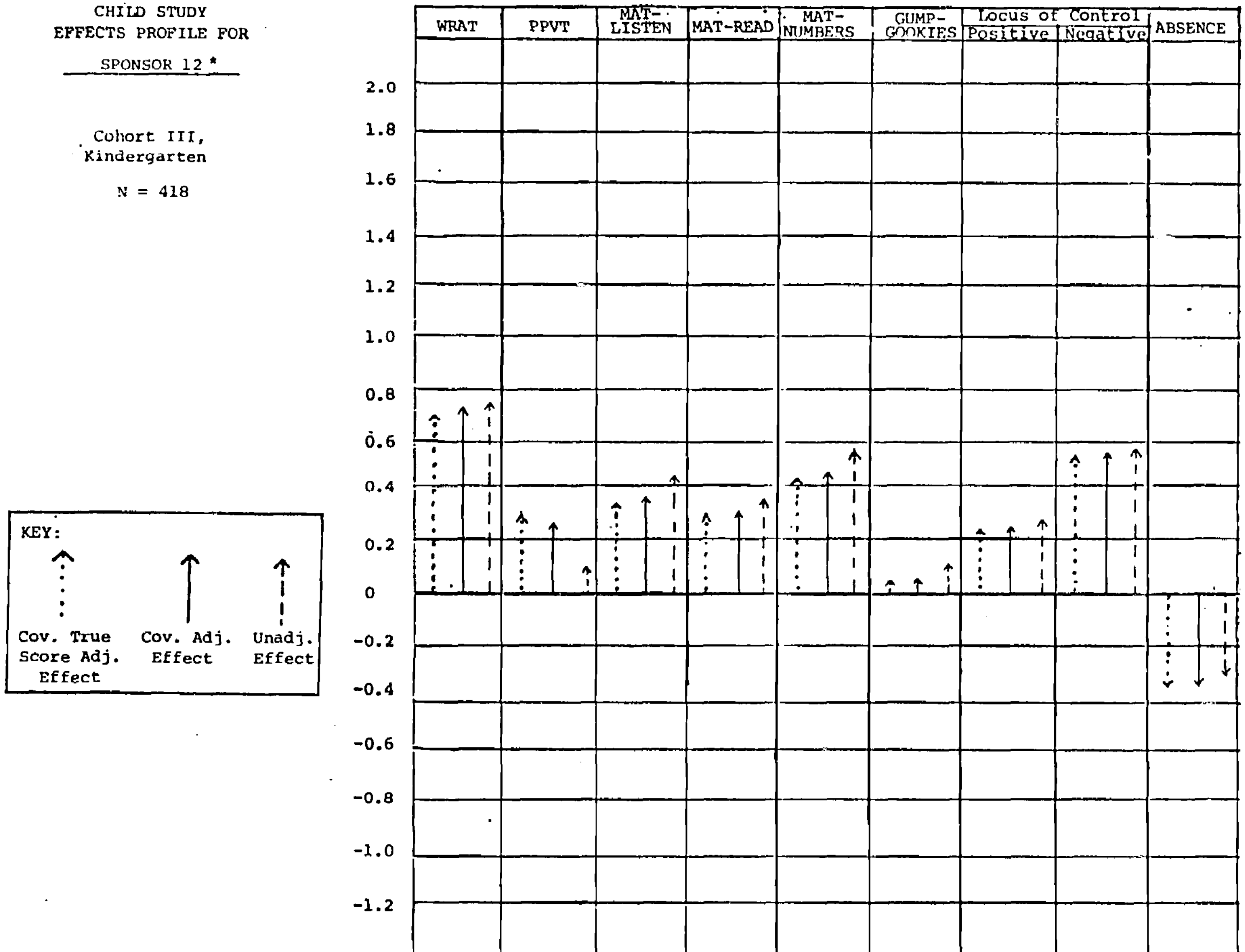
		WRAT	MAT LISTENING	MAT READING	MAT NUMBERS	GUMP-GOOKIES	LOCUS OF CONTROL POSITIVE	LOCUS OF CONTROL NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	10.500	2.265	1.651	2.461	-.299	.522	.674	-2.119
	Unadj.	11.392	3.640	2.317	3.860	1.356	.886	.604	-2.855
* B/S = Effect in Standard Deviations	Adjusted	1.468	.518	.545	.658	-.088	.655	1.223	-.350
	Unadj.	1.593	.833	.765	1.033	.399	1.112	1.094	-.472
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

Figure VII - 38

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 12 *

Cohort III,
Kindergarten

N = 418



		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-COOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	9.157	2.170	2.726	1.603	2.839	.167	.432	..906	-4.167
	Adjusted	9.578	1.906	2.895	1.740	3.000	.246	.459	.916	-4.204
	Unadjusted	9.813	.798	3.426	2.063	3.860	.941	.581	.923	-3.746
*B/S = Effect in Standard Deviations	True Score Adj.	.718	.285	.351	.285	.425	.020	.214	.562	-.351
	Adjusted	.751	.250	.373	.310	.449	.029	.227	.568	-.354
	Unadjusted	.769	.105	.441	.367	.578	.112	.288	.573	-.316
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

and more experienced than the average FT teacher. A somewhat smaller percentage of these FT teachers have obtained advanced credits or degrees, however, and their salaries are slightly lower than average. Approximately 94% of the teachers are White, reflecting the predominantly White student population. The NFT group is similar in each of these respects to the FT group.

The FT teachers report receiving a great deal of Sponsor training in a variety of areas, including the use of structured, sequenced materials, individualization of instruction, and how to work effectively with parents and aides. In these, teachers value using structured learning activities to teach basic skills and involving parents in the education of their children more than do their NFT counterparts. In addition, they visit pupil homes more than the NFT teachers. Finally, the Pittsburgh teachers are somewhat more satisfied than the average FT teacher and perceive themselves as being more faithful to their Sponsor's approach than does any other FT group.

4.9.5 Summary and Discussion

Across all three levels of analysis, the FT group exceeded the NFT group on two achievement outcomes--the MAT arithmetic subtest, and the WRAT--and on the locus of control measures. These contrasts are extremely consistent, despite differences in the geographic distribution of the samples and the characteristics of the pupils served. At least on these outcome variables, there appears to be little variability in the effectiveness of the program, in working with a variety of types of pupils, classes, schools, and communities.

On each of the other achievement and affective variables, however, there are differences in the results across levels of analysis. The similarity of the class and child results, and their dissimilarity with the school level results, may be a function of several things: (1) the relatively high entering achievement of the FT children served at the class and child levels, (2) the overrepresentation of one rural site at these two levels, (3) unmeasured differences in the characteristics of teachers, parents, or classes across levels, or (4) a combination of these. Future analyses will explore these alternative hypotheses.

Overall, it does appear that the University of Pittsburgh program is having measurable impact on both the achievement and motivation of kindergarten children. It must be remembered, however, that this FT group is higher on initial achievement and on mother's education than any other FT group. Once again, we will want to assess the effectiveness of this Sponsor with a variety of types of children, in a variety of environmental contexts, at various stages of child development.

4.10.0 SPONSOR 14: SOUTHWEST EDUCATION DEVELOPMENT LABORATORY (SEDL)

The Language Development (Bilingual) Approach was originally designed as an instructional program for predominantly Spanish-speaking classrooms. The primary emphasis of the approach is on language development; language is seen as the tool for acquiring a variety of skills including non-linguistic skills. Building upon the child's native language and culture, the kindergarten program stresses the development of visual, auditory, and motor skills, as well as thinking, discovery, and English language structures.

4.10.1 School Level FT/NFT Contrasts

The subset of schools which were selected for inclusion in the school level analyses were drawn from three of Sponsor 14's sites. One is a large Northeastern city, another a small Western city, and the third a Southern, rural community. The schools are fairly evenly distributed among these three sites.

The FT schools for this Sponsor serve an extremely disadvantaged group of children, relative to the total group of schools for all Sponsors. The mean adjusted income for this group is far less than the average for all Sponsors. So too, the educational level of the mothers of the children served by these schools is extremely low; the percentage of FT mothers completing high school is 27%. The FT children are primarily from minority groups, with roughly equal proportions of Black and Spanish-surnamed children. Finally, the SEDL FT schools have a lower mean score on the Fall WRAT than any other FT group.

Comparing the FT/NFT schools for this Sponsor, we find that there are several differences between the two groups. Despite the fact that the NFT group is also well below the average, compared to other NFT groups in this sample, it still exceeds the FT group on both SES and entering achievement scores. The NFT group also has a lower mean percentage of minority pupils than the FT group. The mean percentage of minority pupils is 77% for the FT group and 63% for the NFT group.

Figure VII-20 presents the school level FT/NFT contrasts for the SEDL program on the Spring outcome measures. When initial differences are taken into account, there are only two significant contrasts for this Sponsor, both in the affective domain. The FT group scores lower than the NFT group on the Gumpgookies test. On the other hand, the FT group is absent five fewer days than the NFT group, on the average. Given the variability across schools in other outcomes, however, it is also likely that at least some FT schools exceed their NFT counterparts on the MAT reading and arithmetic subtests, as well as on the Locus of Control measures.

4.10.2 Class Level FT/NFT Contrasts

An examination of the distribution of classes by sites reveals that over half of the FT/NFT classes in this group are located in the small Western site, 40% in the Southern site, and very few in the Northeastern site. This geographic distribution differs from that found at the school level, where the schools were more evenly divided among sites.

This change in geographic distribution is paralleled by a change in the demographic makeup of the FT/NFT groups at the class level of analysis. Whereas at the school level the NFT group exceeded the FT group on both SES and entering achievement measures, at the class level this is no longer true. Here, while the NFT group remains slightly higher in mean adjusted income level, it is no different from the FT group in the mean percentage of mothers completing high school. Furthermore, the FT group exceeds the NFT group slightly in entering achievement at this level of analysis.

Finally, there is no change in the mean proportion of minority children in the FT/NFT groups from school to class level. The mean percentage of minority pupils is higher in FT classes (72%) than in NFT classes (53%). However, the percentage of Spanish-speaking children is different. The Western site is the only one in this Sponsor's subset which contains large numbers of Spanish-speaking children in FT classes. Thus, at the class level there is a higher percentage of these children available for

Figure VII - 20

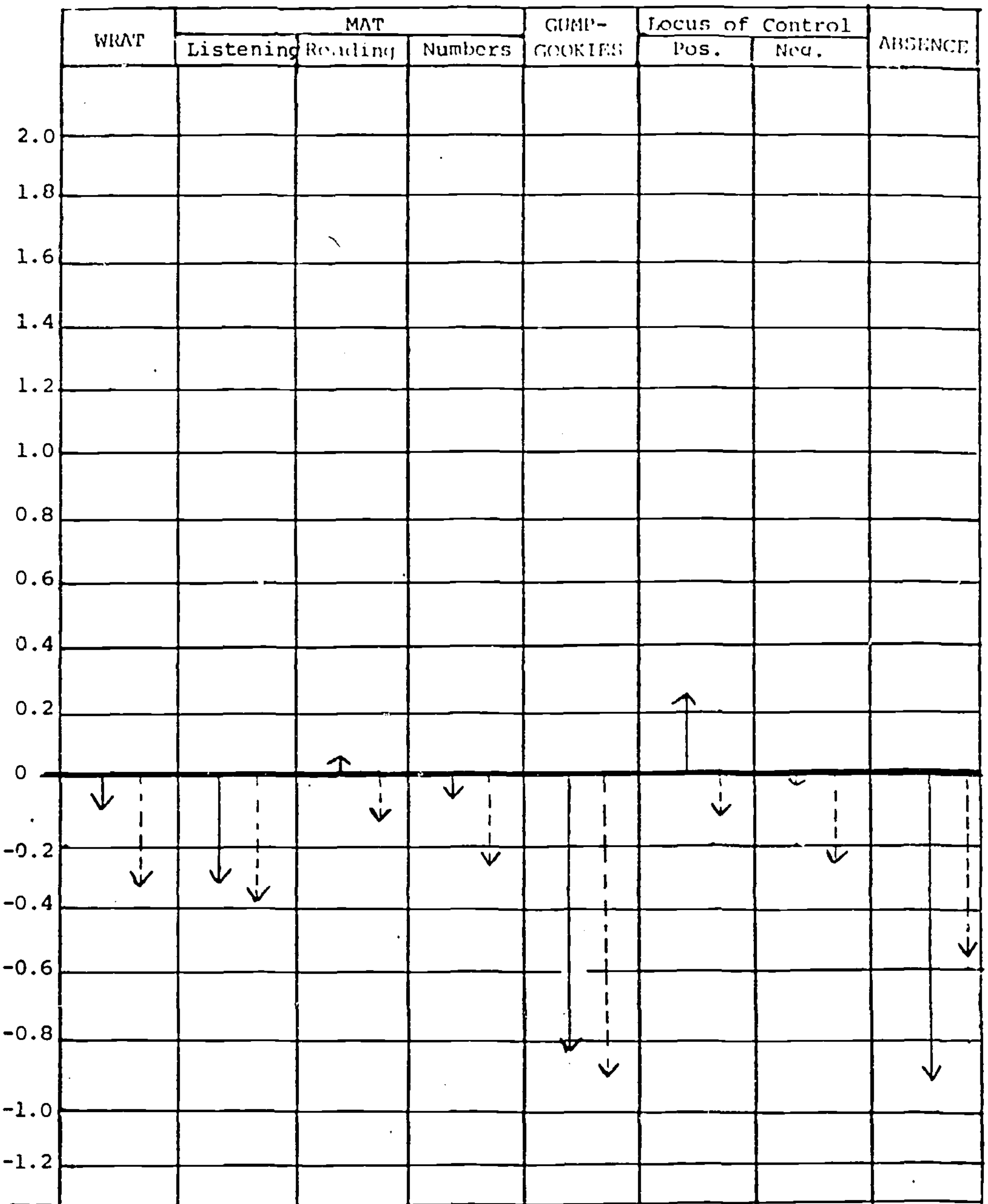
FOLLOW THROUGH EFFECTS PROFILE FOR SPONSOR 14

Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY:

Cov. Adj. Effect — Unadj. Effect



		WRAT	MAT Listening	MAT Reading	MAT Numbers	COOP-COOKIES	Locus of Control Pos.	Locus of Control Neg.	ABSENCE
B = Magnitude of Effect	Adjusted	-0.081	-1.191	0.145	-0.159	-10.222	0.142	-0.007	-5.201
	Unadj.	-0.294	-1.452	-0.276	-0.731	-10.955	-0.052	-0.081	-3.310
SE = Standard Error of B	Adjusted	0.270	1.271	0.863	1.051	4.823	0.218	0.168	2.247
	Unadj.	0.383	1.561	1.070	1.301	5.662	0.270	0.171	2.638
t = B/SE = "Significance" Statistic	Adjusted	-0.300	-0.937	0.168	-0.151	- 2.119	0.653	-0.044	-2.314
	Unadj.	-0.768	-0.930	-0.259	-0.562	- 1.935	-0.192	-0.472	-1.255
S = Standard Deviation		0.884	3.722	2.434	3.104	12.530	0.568	0.352	5.759
*B/S - Effect in Standard Deviations	Adjusted	-0.091	-0.320	0.059	-0.051	- 0.816	0.250	-0.020	-0.903
	Unadj.	-0.333	-0.390	-0.114	-0.236	- 0.874	-0.091	-0.229	-0.575
N = Number of Schools in Computation	FT	9	9	9	9	9	9	9	9
	NFT	8	8	8	8	8	8	8	8

analysis than at the school level. The class level, therefore, provides a better opportunity for examining the impact of the program upon the children for whom it was originally intended.

The class level FT/NFT contrasts in achievement are markedly different from those found at the school level (see Figure VII-39). At the class level, the FT group exceeds the NFT group significantly on all achievement outcomes, with the contrasts being especially large on the MAT Listening and Reading subtests. As in the school level analyses, the FT group also exceeds the NFT group substantially in attendance, but scores lower than the NFT group on the Gumpgookies test. Finally, the FT group scores lower than the NFT group on the Locus of Control (positive) measure, a finding which is inconsistent with the positive trend found at the school level.

4.10.3 Child Level FT/NFT Contrasts

Over 70% of the FT children in the child level analysis were drawn from the Southern site, a much higher percentage than in either the school or class analyses. On two demographic characteristics, however, the class and child FT/NFT groups are similar. At the child level of analysis, the FT group is lower than the NFT group on adjusted income level and higher than the NFT group on the Fall WRAT, differences which parallel the FT/NFT differences at class level. On the other hand, the mean percentage of FT mothers completing high school slightly exceeds the mean percentage of NFT mothers completing high school, whereas the two groups were the same at class level.

The subset of children chosen for these analyses, however, differs markedly from the subsets of schools and classes in ethnic composition. At the school level of analysis roughly 30% of the FT children and 20% of the NFT children were Spanish-speaking. So too, the subset of classes contained a number of Spanish-speaking children, as mentioned above. The subset of children meeting the criteria for inclusion in the child level analyses, however, included only Black and native English-speaking White children. At the child level of analysis, the percent Black FT children was 65%, and the percent Black NFT children was 57%.

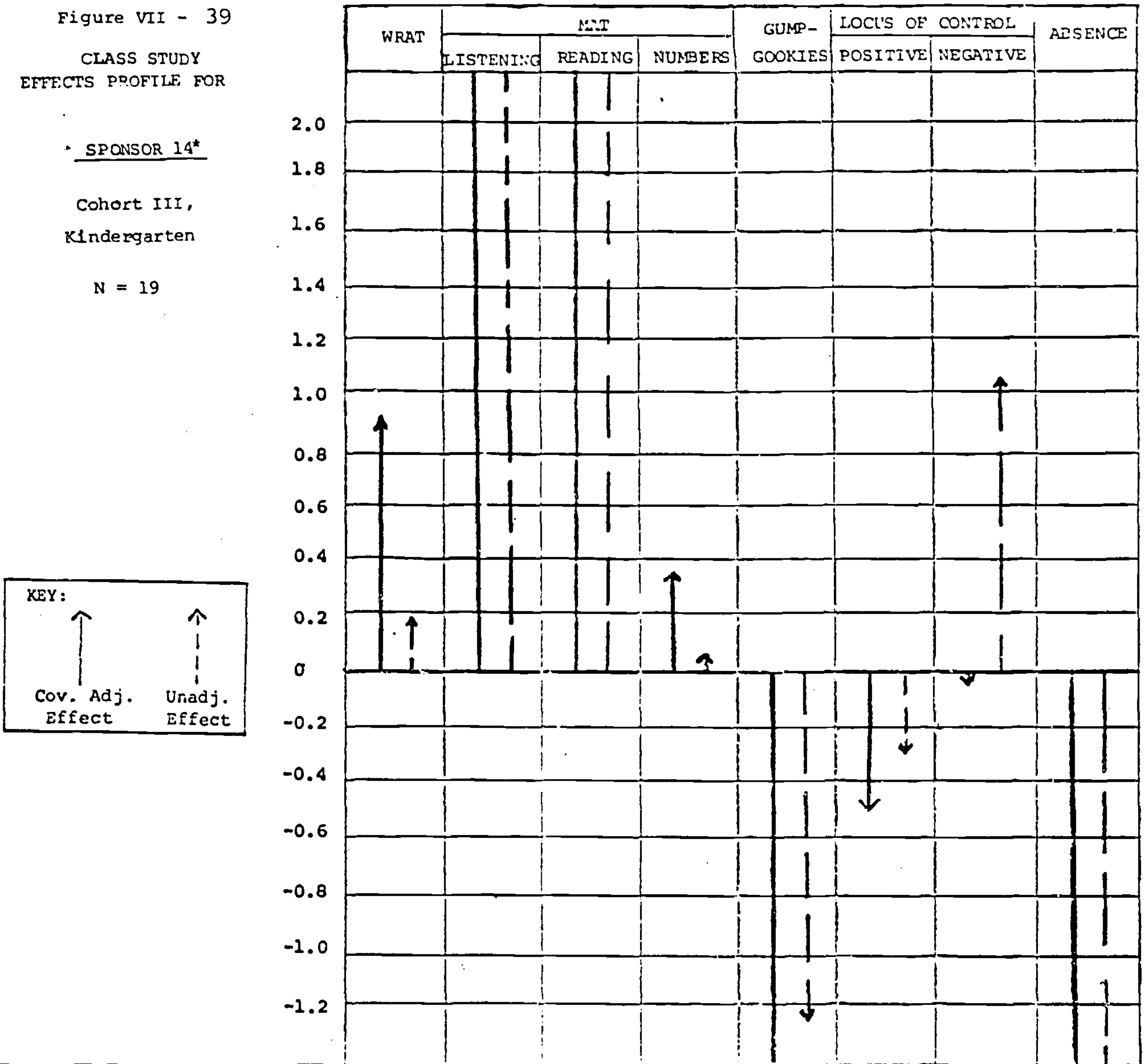
Figure VII - 39

CLASS STUDY
EFFECTS PROFILE FOR

* SPONSOR 14*

Cohort III,
Kindergarten

N = 19



		WRAT	LISTENING	READING	NUMBERS	GUMP-GOOKIES	POSITIVE	NEGATIVE	ABSENCE
B = Magnitude of Effect	Adjusted	6.494	9.635	14.102	1.370	-6.413	-.417	-.011	-13.582
	Unadj.	1.421	11.467	15.512	.037	-4.181	-.198	.555	-8.656
*B/S = Effect in Standard Deviations	Adjusted	.908	2.205	4.657	.366	-1.887	-.523	-.020	-2.245
	Unadj.	.199	2.624	5.123	.010	-1.230	-.249	1.007	-1.431
S = Standard Deviation		7.152	4.370	3.028	3.739	3.399	.797	.551	6.049

Despite geographic, demographic, and ethnic/language differences, the child level findings parallel the school level findings in all but one respect. At the child level, as in the class level analyses, the FT group exceeds the NFT group significantly on the MAT Reading subtest. In addition, the FT group exceeds the NFT group on the PPVT as well. (See Fig. VII-40.)

It appears that the SEDL program may not only be successful in developing the reading skills of Spanish-speaking children, but it may also be useful with other types of children as well.

4.10.4 Selected Teacher Data

The FT teachers in the bilingual approach are somewhat younger and less experienced than the average FT teacher. They are above average in educational attainment, however, with 82% having obtained advanced credits or degrees. Approximately 70% of the FT teachers are White.

The NFT teachers for Sponsor 14 are older and more experienced than the FT teachers. Approximately 90% have obtained advanced credits or degrees, and all are White.

The teachers in this program report receiving relatively little Sponsor training, compared to other FT teachers. The training they do receive is primarily in the use of small groups and sequenced materials to structure the learning environment. The FT teachers for this Sponsor place greater value on the development of respect for the rights of others and pupil cooperation than do their NFT counterparts or other FT teachers. They also place great value on the structured approach to teaching basic skills, both relative to their NFT group and other FT teachers. Finally, they make a great many visits to pupils' homes.

4.10.5 Summary and Discussion

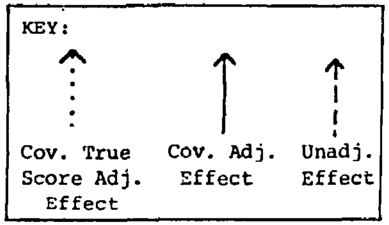
In the area of achievement, the SEDL program appears to be having some success in developing listening and reading skills. While this Sponsor's impact does not appear to be limited to Spanish-speaking children, it does vary with the communities in which the Sponsor operates and with

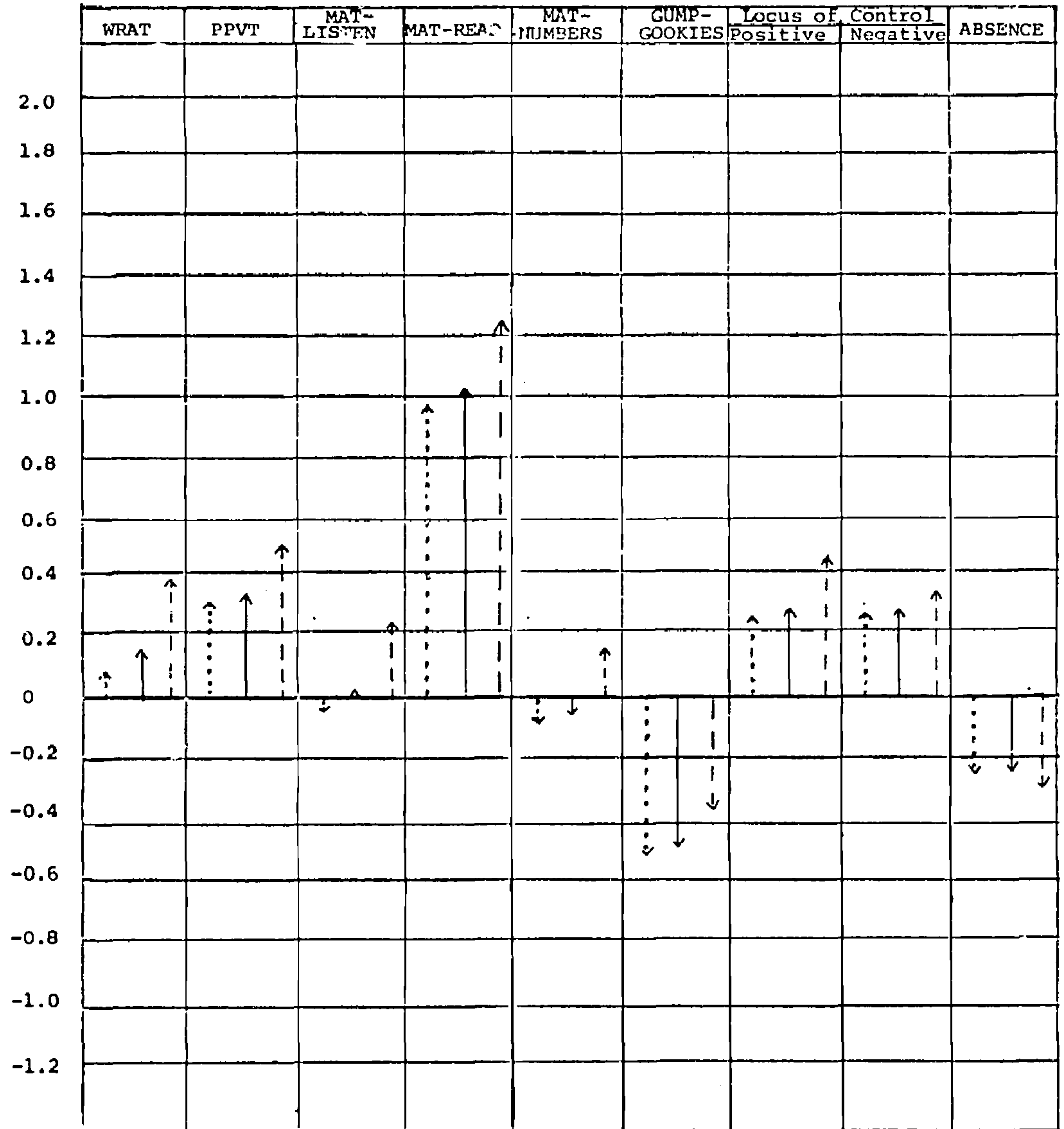
Figure VII - 40

CHILD STUDY
EFFECTS PROFILE FOR
SPONSOR 14*

Cohort III,
Kindergarten

N = 136





		WRAT	PPVT	MAT-LISTEN	MAT-READ	MAT-NUMBERS	GUMP-GOOKIES	Locus of Control Positive	Locus of Control Negative	ABSENCE
B = Magnitude of Effect	True Score Adj.	1.197	2.399	-.260	5.509	-.600	-4.199	.523	.414	-2.719
	Adjusted	2.006	2.636	.065	5.773	-.292	-4.048	.573	.432	-2.790
	Unadjusted	5.069	4.108	1.673	7.032	1.069	-3.144	.876	.552	-3.490
*B/S = Effect in Standard Deviations	True Score Adj.	.094	.315	-.033	.980	-.090	-.501	.259	.257	-.229
	Adjusted	.157	.346	.008	1.027	-.044	-.483	.284	.268	-.235
	Unadjusted	.397	.539	.215	1.250	.160	-.375	.434	.343	-.294
S = Standard Deviation		12.754	7.615	7.769	5.618	6.674	8.381	2.016	1.611	11.869

the characteristics of the pupils served. Differences in the characteristics of the samples at the various levels of analysis suggest that this Sponsor may be less effective with children in the large Northeastern cities and with children whose entering achievement and mother's education is extremely low. Further exploration into the effectiveness of the SEDL program in producing achievement results with both Spanish-speaking and native Engligh-speaking children, in a variety of community settings, is needed.

In the affective domain, the SEDL program appears to be having a positive effect on children's Locus of Control. It may be that the use of positive reinforcement techniques and frequent adult feedback, which are basic strategies of the SEDL approach, results in FT children learning that their actions lead to positive events in the real world. These contrasts are small, however, and they also vary with the subset of schools, classes, and children analyzed.

The attendance data are more consistent. FT children in the SEDL program are found to attend school more often than their NFT counterparts at each level of analysis studied. As has been discussed elsewhere, this increase in attendance may mean one of at least three things: 1) FT children are healthier; 2) FT children enjoy school more and so are more eager to attend; and 3) FT parents are more apt to send their children to school. However, the fact that FT children for this Sponsor do not score higher than NFT children on the Gumpgookies test, which is designed to measure achievement motivation and school enjoyment, makes the second alternative seem unlikely in this case. Whatever the reason, the increase in attendance is encouraging. For educators, regular attendance means less interruption of the learning sequence and more opportunity for instruction. For administrators, regular attendance means efficiency and economy.

Finally, we will discuss the Gumpgookies contrasts. Given the very poor families from which these FT children come, it may be that the relatively low Gumpgookies scores for this group indicates that achievement motivation and mastery are very low on the hierarchy of needs. On

the other hand, the heavy emphasis on basic skills and the highly structured learning environment advocated by this Sponsor, and valued by these teachers, may be having a negative effect on children's enjoyment of school and discouraging independent, purposive behavior. Once again, future analyses will systematically explore these alternative hypotheses. They will also allow us to examine children's growth patterns over time.

5.0 COHORT I AND MULTIPLE COHORT STUDIES

5.1.0 INTRODUCTION, COHORT I STUDY

The major questions that motivate the FT national evaluation are longitudinal, as we have already suggested, but our Cohort III findings are as yet only cross-sectional. We have outlined the severe limitations of the data from Cohorts I and II: becuase of these limitations we have chosen to rely almost exclusively on Cohort III data for this report's analyses. With all of the drawbacks in Cohorts I and II, however, we can perhaps draw from the early data some tentative longitudinal context for our present cross-sectional findings.

5.2.0 METHOD

5.2.1 Analytic Subset

Accordingly, we now present a study of FT/NFT contrasts in a restricted sample of Cohort I children who entered FT in 1969 as first graders. These children completed Head Start in the Spring of 1969, began their elementary education the following Fall without a kindergarten experience, and continued through the third grade as members of the FT program. They come primarily from the Southern sections of the nation where kindergartens are not available. They are also the first group of children with whom each of the Sponsors were involved at each successive grade level, after the original implementation year of 1968-69. Thus, they represent a unique group of children, interacting with a unique aspect of the Sponsors' programs.

The analysis presented below, while tentative, is an attempt to examine the first group of FT graduates. Later analyses along similar lines will shed some light on the longitudinal questions generated by FT. Does FT continue to be beneficial to children throughout the three or four years during which it is designed to intervene in their lives? Have the FT Sponsors succeeded in overcoming the damping-out effects that have been observed in much past research on the long-term consequences of preschool compensatory intervention? It is still much too early to ask our evaluative data for answers to these questions. This initial three-year longitudinal study indicates the manner in which

we shall approach them when the appropriate data become available. While it would be grossly unreasonable to judge FT by these early indications, they may suggest trends which we shall watch for in later analyses.

Table VII-11 displays the population for this three-year longitudinal study which consists of 40 schools, distributed among 6 Sponsors.

As in the kindergarten studies, we have contrasted FT and NFT groups at school level, adjusting postscore differences to compensate for initial mismatch. Five of the six Sponsors are common to the two sets of analyses: Sponsor 6 is not included in the kindergarten studies. Here, our outcome measures are the three subtests of the third grade (1972) MAT: reading, arithmetic, and spelling. No psychometrically-equivalent pretest was administered in 1969 at the beginning of first grade; we therefore used the results of the Pre-School Inventory (PSI) and Wide Range Achievement Test (WRAT) as (surrogate) measures of entry-level achievement for purposes of covariate adjustment. Average months of preschool experience entered the analysis as a third covariable.

The children included in the computations of school scores were limited to those who entered the program in the Fall of 1969, remained with the same Sponsor and school through the third grade, and were tested both at the beginning of the first grade and the end of the third grade. As one might expect, these stringent conditions reduced the analysis population substantially. Of the 9,879 children listed on the first grade roster in the Fall of 1969, only 4,316 received either the first or the second test. Of these, moreover, only 1,216 children were tested both times and were therefore eligible for inclusion in this study.

5. 2.2 Design

The analytic model for this analysis takes the same form as that for the school level kindergarten studies reported in Section 1.2 of this Chapter. Tables VII-12 and VII-13 display the predictor coding schemes for the factorial analysis and the nested analysis, respectively.

Before going on to display results, let us reiterate some of the numerous ways in which this study is not comparable to the one-year effects study:

Table VII - 11

Distribution of Schools in the Three-Year Longitudinal Effects Study Population by Sponsor by FT/NFT

Sponsor	Follow Through	Non-Follow Through	Total
5	4	2	6
6	8	3	11
7	3	2	5
9	6	3	9
11	1	2	3
12	3	3	6
All Sponsors	25	15	40

TABLE VII - 12

Contrast Coding Scheme for the Factorial Analysis of Three-Year Effects

Sponsor	FT/NFT Status	5 Contrasts Among Individual Sponsors					Main FT/NFT Effect	Interactions of Sponsor Contrasts with the Main Effect				
		4	5	6	7	8	9	10	11	12	13	14
5	FT	1	0	0	0	0	.5	.5	0	0	0	0
	NFT	1	0	0	0	0	-.5	-.5	0	0	0	0
6	FT	0	1	0	0	0	.5	0	.5	0	0	0
	NFT	0	1	0	0	0	-.5	0	-.5	0	0	0
7	FT	0	0	1	0	0	.5	0	0	.5	0	0
	NFT	0	0	1	0	0	-.5	0	0	-.5	0	0
9	FT	0	0	0	1	0	.5	0	0	0	.5	0
	NFT	0	0	0	1	0	-.5	0	0	0	-.5	0
11	FT	0	0	0	0	1	.5	0	0	0	0	.5
	NFT	0	0	0	0	1	-.5	0	0	0	0	-.5
12	FT	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
	NFT	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5

TABLE VII - 13

Contrast Coding Scheme for the Nested Analysis for Three-Year Effects

Sponsor	FT/NFT Status	6 Follow Through Effects Within Sponsors						5 Contrasts Among Sets of Sponsors				
		4	5	6	7	8	9	10	11	12	13	14
5	FT	.5	0	0	0	0	0	1	0	-2	0	2
	NFT	-.5	0	0	0	0	0	1	0	-2	0	2
6	FT	0	.5	0	0	0	0	1	1	1	1	-1
	NFT	0	-.5	0	0	0	0	1	1	1	1	-1
7	FT	0	0	.5	0	0	0	1	-1	1	-1	-1
	NFT	0	0	-.5	0	0	0	1	-1	1	-1	-1
9	FT	0	0	0	.5	0	0	-1	0	-2	0	-2
	NFT	0	0	0	-.5	0	0	-1	0	-2	0	-2
11	FT	0	0	0	0	.5	0	-1	1	1	-1	1
	NFT	0	0	0	0	-.5	0	-1	1	1	-1	1
12	FT	0	0	0	0	0	.5	-1	-1	1	1	1
	NFT	0	0	0	0	0	-.5	-1	-1	1	1	1

- Population: No children are common to the two studies. Five of the six Sponsors from this study are also included among the ten Sponsors in the one-year study. This study involves only children traced over a three-year period and therefore reflects a bias for geographical stability (a covariable in the one-year study). Data from Cohort I doubtlessly reflect the problems (and Hawthorne benefits) of start-up more than data from the later, more experienced Cohort III.

- Variables: In the Cohort I longitudinal study there is no true pretest measure for any of the criterion variables in the sense that Fall WRAT is a pretest for Spring WRAT in the one-year study. The third grade MAT subtests are analogous to the achievement tests that we used as criteria in the one-year study.

Any similarities in the results of the two studies must therefore reflect either coincidences or truly pervasive patterns of the sort that our cross-validation strategy is designed to detect.

5.3.0 RESULTS

With all these caveats, we now present in Table VII-14 the regression statistics for the three-year effects analysis. As in the one-year study, the covariables account for about half of each criterion's variance, and the FT and Sponsor predictors together account for roughly another quarter. In this six-Sponsor set, main effects for the three MAT subtests do not stand very substantially above the noise, which is considerable. The F statistics for Sponsor effects indicate that a significant proportion of variance is accounted for ($P < .05$) in the spelling outcome; the F statistics also indicate significant effects for Sponsor x FT interactions for both the reading and spelling outcomes. FT related factors account for substantially less variance in the MAT arithmetic score than in the reading and spelling scores. For reading and spelling, the message of the data seems to be much the same as in the one-year analysis: Sponsors have widely varying effects.

Given the non-probabilistic nature of the FT quasi-experiment, the substantial differences between the populations and designs of the one-year and three-year studies, and the small "sample" size in this three-year study, we should probably be less concerned about the

TABLE VII - 14

Partition of Variance for the Three-Year Effects Study

Statistic	Criterion: MAT Reading	Arithmetic	Spelling	d.f.
$R^2_{Y \cdot A}$	.55064	.5373	.4738	
$R^2_{Y \cdot AD}$	.72324	.6689	.6363	
$R^2_{Y \cdot ABC} = R^2_{Y \cdot AE}$	.62230	.5746	.5706	
$R^2_{Y \cdot ABD}$	.82129	.7244	.7578	
$R^2_{Y \cdot ACD}$	.73249	.6900	.6392	
$R^2_{Y \cdot ABCD}$	.82450	.7363	.7578	
F_c (FT Main Effect) $= \dfrac{(R^2_{Y \cdot ABCD} - R^2_{Y \cdot ABD})/1}{(1 - R^2_{Y \cdot ABCD})/(N - a - 2s)}$	0.457	1.70	0.01	1,25
F_e (Sponsor Effects) $= \dfrac{(R^2_{Y \cdot ABCD} - R^2_{Y \cdot AD})/s}{(1 - R^2_{Y \cdot ABCD})/(N - a - 2s)}$	2.40	1.60	2.88*	6.25
F_b (Sponsor X FT Interaction Effects) $= \dfrac{(R^2_{Y \cdot ABCD} - R^2_{Y \cdot ACD})/(s - 1)}{(1 - R^2_{Y \cdot ABCD})/(N - a - 2s)}$	2.62*	1.32	3.38*	5,25

KEY:
N = 40 Schools
s = 6 Sponsors
a = 3 Covariates

Predictor Set/Composition
A 3 Covariates
C 1 Main FT Effect
E 6 Sponsor Effects
B 5 Interactions (Sponsor X FT Effects)
D 5 Sponsor Contrasts

* $p \leq .05$

statistical significance of the three year results than about qualitative comparisons between the two studies. Figures VII-41 through VII-43 display the main FT and Sponsor effects for the three-year longitudinal study. Sponsor 12 consistently has the highest positive effects on all three measures; in the one-year study as shown in Figure VII-10. Sponsor 12 has sizeable effects on the arithmetic test but not on the reading and sounds tests. Sponsors 7 and 9 have only negative effects at the end of third grade, in marked contrast to their positive patterns at the end of kindergarten. Covariance adjustment generally enhanced the effects of Sponsors 7 and 9 in the kindergarten study; perhaps a more adequate covariate set would have made the picture more favorable in the three-year study as well. An alternative explanation might be that these Sponsors have positive effects on kindergarteners which are lost by the time these children finish the third grade. A third explanation might be that the population differences between the studies swamped all other influences on the patterns. It is still too early to account for the results; reliable replications of this study in later cohorts will give us a better basis for confident conclusions.

5.4.0 DISCUSSION

One thing we can say with confidence about this longitudinal study is that three years of FT experience have not homogenized the Sponsors. Even within the limited scope of a six-Sponsor study, Sponsor effects remain widely variable. To say more, with reasonable confidence in our interpretations, we shall have to wait for data from the heavily-tested Cohort III when it completes its FT experience in 1975.

These unimpressive three year findings suggest the possibility that Sponsor effects in the early years of the planned variation program were inhibited by early implementation problems which may have persisted in the Sponsors' first dealings with first and second grade curricula. As Sponsors gain experience, and as schools and teachers become adept at working with Sponsors' models, perhaps both the positive and negative consequences of novelty will wear off, permitting the long-term, replicable aspects of Sponsor performance to show through.

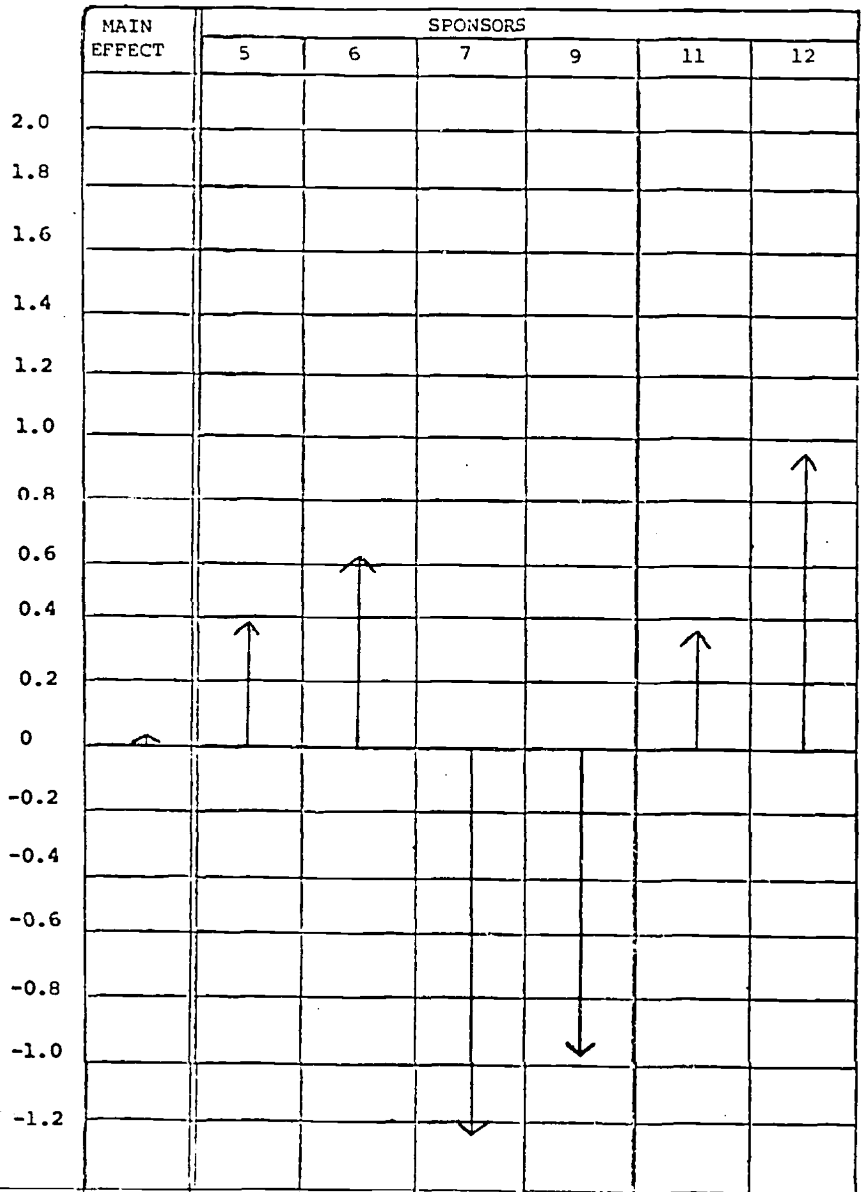
Figure VII - 41

EFFECTS PROFILE FOR:

Three Year Longitudinal Study

MAT: Spelling

S.D. = 6.65



	MAIN EFFECT	5	6	7	9	11	12
B = Magnitude of Effect	0.16	2.61	4.15	-8.18	-6.42	2.44	6.39
SE_B = Standard Error of B	1.87	4.13	3.26	4.36	3.48	5.01	3.61
Significance Statistic	0.09	0.63	1.27	-1.88	-1.84	0.49	1.77
B/S.D. = Effect in Standard Deviations	0.024	0.39	0.62	-1.23	-0.96	0.37	0.96
N = Number of Schools	40	6	11	5	9	3	6

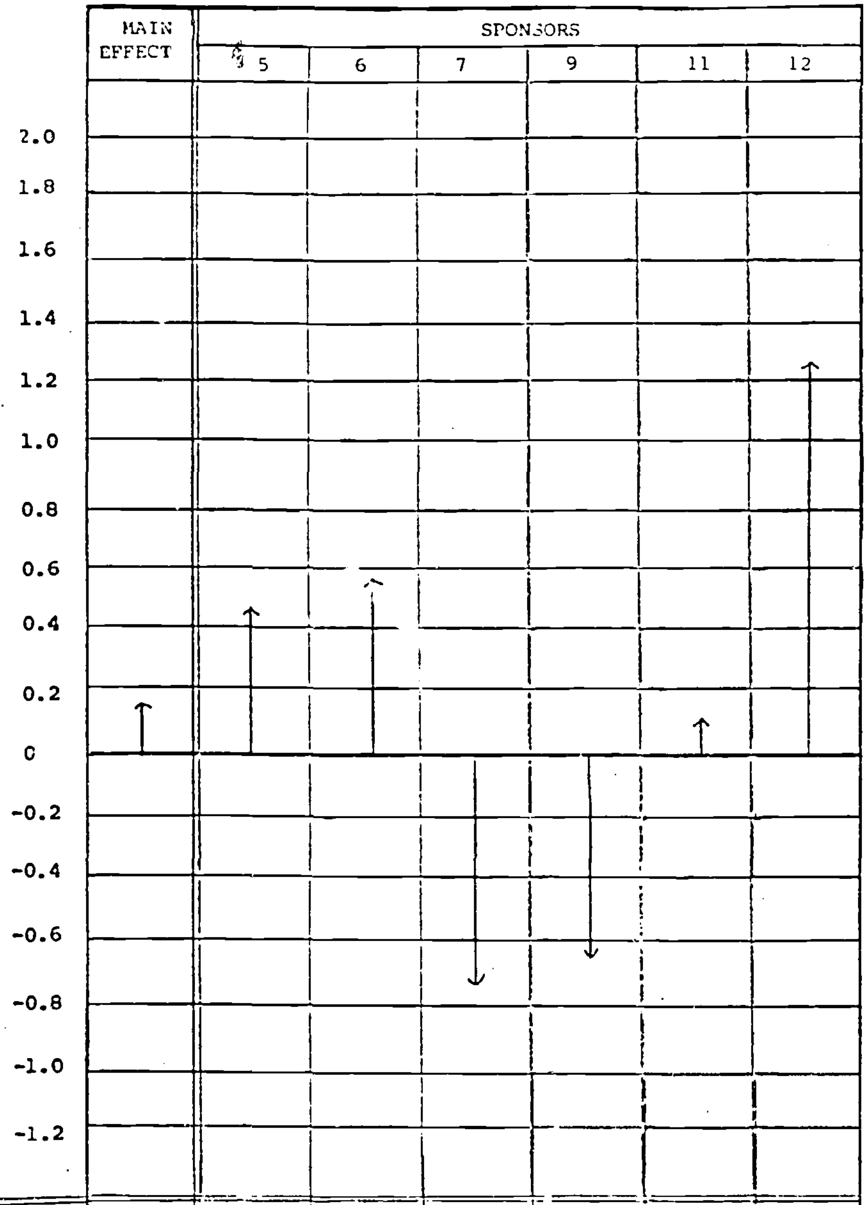
Figure VII - 42

EFFECTS PROFILE FOR:

Three Year Longitudinal Study

MAT: Reading

S.D. = 11.85



	MAIN EFFECT	5	6	7	9	11	12
B = Magnitude of Effect	1.93	5.39	6.82	-8.95	-7.60	1.08	14.81
SE_B = Standard Error of B	2.84	6.27	4.96	6.61	5.28	7.61	5.48
Significance Statistic	0.68	0.86	1.38	-1.35	-1.44	0.14	2.70
B/S.D. = Effect in Standard Deviations	0.16	0.46	0.58	-0.76	-0.64	0.11	1.25
N = Number of Schools	40	6	11	5	9	3	6

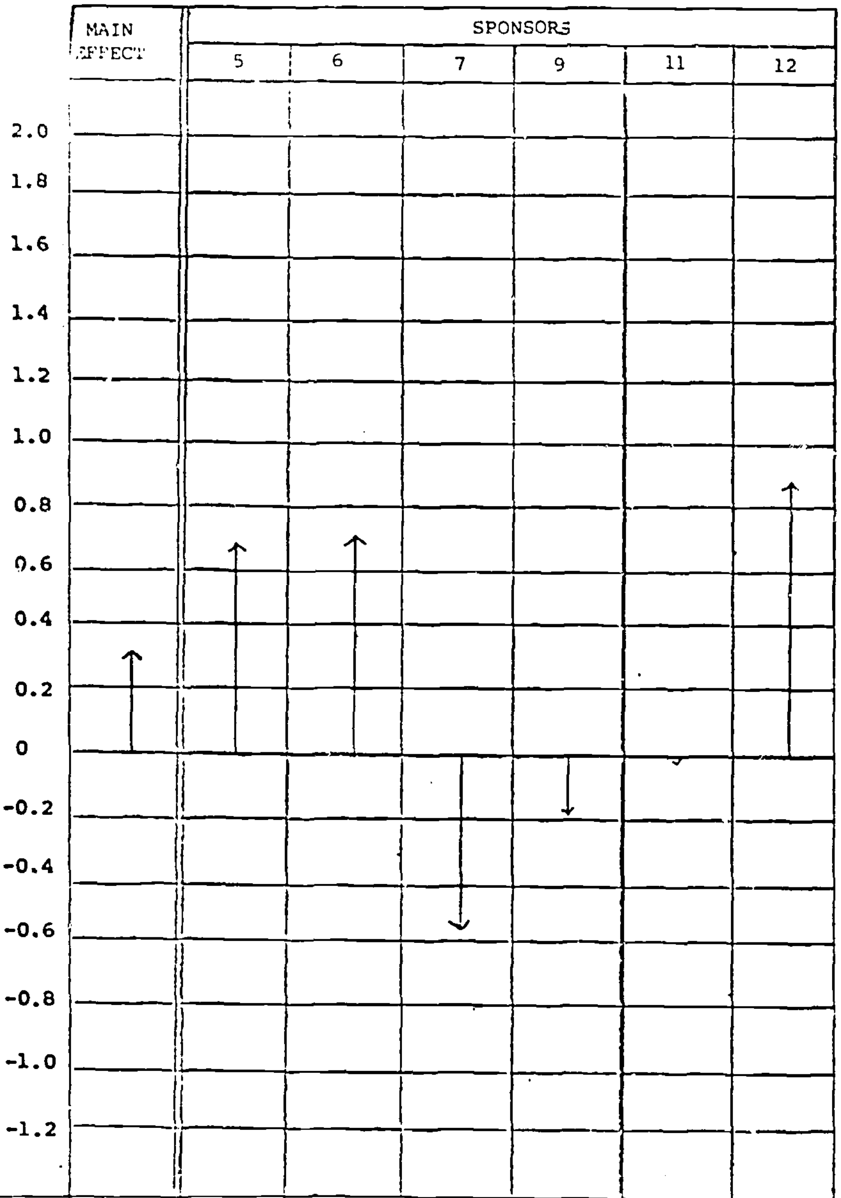
Figure VII - 43

[illegible]

<u>Three Year Longitudinal Study</u>

MAT: Arithmetic

S.D. = 13.1



	Main Effect	5	6	7	9	11	12
B = Magnitude of Effect	4.05	8.97	9.19	-7.52	-2.53	-0.07	11.32
SE_B = Standard Error of B	3.83	8.47	6.67	8.92	7.12	10.27	7.40
Significance Statistic	1.06	1.06	1.38	-0.84	-0.36	-0.01	1.53
B/S.D. = Effect in Standard Deviations	0.31	0.68	0.70	-0.57	-0.19	-0.005	0.86
N = Number of Schools	40	6	11	5	9	3	6

5.5.0 INTRODUCTION, MULTIPLE COHORT STUDY

The data thus far provide very little opportunity to investigate the question of Sponsor maturation. By way of preparing for better data to come, we have made a crude initial study of six Sponsors working with kindergarten children from two different Cohorts.

5.6.0 METHOD

5.6.1 Analytic Subset

We examined mean WRAT scores, Fall and Spring, in 24 schools which participated in both Cohort I (1969-70) and Cohort III (1971-72) kindergarten programs with the same Sponsor. Table VII-15 shows the distribution of these schools by Sponsor and FT/NFT. With such a small data set for analysis, we faced an even more unstable situation than the 40-school, three-year effects analysis. The analytic model for this analysis should theoretically extend to triple interactions of Sponsor FT/NFT, and cohort membership, requiring even larger numbers of predictors and reducing still further the number of degrees of freedom available to lend the analysis stability and sensitivity. Only the main effect studies, however, are somewhat indicative. Finally, recent studies of testing schedules have demonstrated that Cohort I was tested systematically later in the Fall than Cohort III; we discuss the apparent effects of these delays in Section 5.8.0. We therefore present the model and results of the multiple cohort study, not so much for the sake of the results but rather to foreshadow more meaningful analyses of similar forms in later reports.

5.6.2 Design

We have subjected each variable of the pupil data, aggregated to the school level to two analyses: analysis A, a study of school variance and analysis B, a study of trend between cohorts.

Analysis A is a between-schools analysis summing across time points and ignoring cohort differences. It asks:

- Are there overall covariance adjusted Sponsor differences?
- Are there overall covariance adjusted FT/NFT differences?

Table VII - 15

Distribution of Schools in the Multiple-Cohort Study Population by Sponsor and by Follow Through Participation Status

Sponsor	Follow Through	Non-Follow Through	Total
2	3	2	5
3	2	1	3
5	3	1	4
8	3	1	4
11	2	2	4
13	2	2	4
All Sponsors	15	9	24

- Are there covariance adjusted Sponsor by FT/NFT interactions?

The model for analysis A takes the following form:

$$\hat{Y} = Y_O + B_G X_G + B_H X_H + B_I X_I + B_{HxI} X_{HxI}$$

Notation for this model is:

$\hat{Y}$ is the estimated value of the criterion variable (Spring WRAT score)

Y_O is the Y intercept

X_G is the covariate (Fall WRAT)

X_i $(i \neq G)$ is the effects-coded parameter for effect i, and

B_i is the raw score regression weight for effect i.

The predictor sets G, H, I and HxI are as follows:

G is the covariate score

H is the Sponsor effect

I is the FT/NFT effect, and

HxI is the Sponsor by FT/NFT interaction.

Table VII-16 displays the coding scheme which defines sets H, I, and HxI.

The model for analysis B takes the following form:

$$\hat{Y} = Y_O + B_G X_G + B_K X_K + B_{HxK} X_{HxK} + B_{IxK} X_{IxK} + B_{HxIxK} X_{HxIxK}$$

The notations Y_O, B, X, G, H, and I are the same in both models. Set K denotes an effects-coded variable embodying cohort membership; KxH, KxI, KxHxI represent the cohort effects interacting with Sponsor, FT/NFT and Sponsor by FT/NFT. This model ignores Sponsor, FT/NFT and Sponsor x FT/NFT effects.

Table VII-17 displays the coding scheme which defines sets K, HxK, IxK, HxIxK. In this table a set of variables J, embodying the variation of schools across Sponsor FT/NFT combinations, is included. This set is used to obtain the total sum of squares for schools eliminating the Y intercept and ignoring the effects of the model; $R^2_{Y \cdot J}$ yields the total sum of squares for the regression analysis, and $R^2_{Y \cdot G,J} - R^2_{Y \cdot G}$ the total sum of squares for the analysis of covariance.

Table VII-18 presents the results of the analyses for the two models.

Table VII - 16

Contrast Coding Scheme for the Sponsor by FT/NFT Interaction Analysis (A) of the Multiple Cohort Study

			Predictor Set H: 5 Contrasts Among Individual Sponsors					Predictor Set I: Treatment	Predictor Set H x I: 5 Contrasts for Sponsor by Treatment				
Sponsor	School	Time	2	3	4	5	6	7	8	9	10	11	12
02	1	1	1	0	0	0	0	.5	.5	0	0	0	0
		2	1	0	0	0	0	.5	.5	0	0	0	0
	2	1	1	0	0	0	0	.5	.5	0	0	0	0
		2	1	0	0	0	0	.5	.5	0	0	0	0
	3	1	1	0	0	0	0	.5	.5	0	0	0	0
		2	1	0	0	0	0	.5	.5	0	0	0	0
	4	1	1	0	0	0	0	-.5	-.5	0	0	0	0
		2	1	0	0	0	0	-.5	-.5	0	0	0	0
	5	1	1	0	0	0	0	-.5	-.5	0	0	0	0
		2	1	0	0	0	0	-.5	-.5	0	0	0	0
03	6	1	0	1	0	0	0	.5	0	.5	0	0	0
		2	0	1	0	0	0	.5	0	.5	0	0	0
	7	1	0	1	0	0	0	.5	0	.5	0	0	0
		2	0	1	0	0	0	.5	0	.5	0	0	0
	8	1	0	1	0	0	0	-.5	0	-.5	0	0	0
		2	0	1	0	0	0	-.5	0	-.5	0	0	0
05	9	1	0	0	1	0	0	.5	0	0	.5	0	0
		2	0	0	1	0	0	.5	0	0	.5	0	0
	10	1	0	0	1	0	0	.5	0	0	.5	0	0
		2	0	0	1	0	0	.5	0	0	.5	0	0
	11	1	0	0	1	0	0	.5	0	0	.5	0	0
		2	0	0	1	0	0	.5	0	0	.5	0	0
	12	1	0	0	1	0	0	-.5	0	0	-.5	0	0
		2	0	0	1	0	0	-.5	0	0	-.5	0	0
08	13	1	0	0	0	1	0	.5	0	0	0	.5	0
		2	0	0	0	1	0	.5	0	0	0	.5	0
	14	1	0	0	0	1	0	.5	0	0	0	.5	0
		2	0	0	0	1	0	.5	0	0	0	.5	0
	15	1	0	0	0	1	0	.5	0	0	0	.5	0
		2	0	0	0	1	0	.5	0	0	0	.5	0
	16	1	0	0	0	1	0	-.5	0	0	0	-.5	0
		2	0	0	0	1	0	-.5	0	0	0	-.5	0
11	17	1	0	0	0	0	1	.5	0	0	0	0	.5
		2	0	0	0	0	1	.5	0	0	0	0	.5
	18	1	0	0	0	0	1	.5	0	0	0	0	.5
		2	0	0	0	0	1	.5	0	0	0	0	.5
	19	1	0	0	0	0	1	-.5	0	0	0	0	-.5
		2	0	0	0	0	1	-.5	0	0	0	0	-.5
	20	1	0	0	0	0	1	-.5	0	0	0	0	-.5
		2	0	0	0	0	1	-.5	0	0	0	0	-.5
13	21	1	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
		2	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
	22	1	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
		2	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
	23	1	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5
		2	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5
	24	1	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5
		2	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5

Table VII - 17

Contrast Coding Scheme for the Sponsor by FT/NFT by Time Analysis (B) of the Multiple Cohort Study

Sponsor	School	Time	Predictor Set J: 23 Contrasts Among 24 Schools																							Predictor Set K: Time	Predictor Set H x K: 5 Contrasts for Sponsor by Time					Predictor Set I x K: Treatment by Time	Predictor Set H x I x K: 5 Contrasts for Sponsor by Treatment by Time				
			2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
02	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	-.5	-.5	0	0	0	0
		2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	.5	.5	0	0	0	0
	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	-.5	-.5	0	0	0	0
		2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	.5	.5	0	0	0	0
	3	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	-.5	-.5	0	0	0	0
		2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	.5	.5	0	0	0	0
	4	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	.5	.5	0	0	0	0
		2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	-.5	-.5	0	0	0	0
	5	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	0	0	.5	.5	0	0	0	0
		2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	-.5	-.5	0	0	0	0
03	6	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-1	0	0	0	-.5	0	-.5	0	0	0
		2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	.5	0	.5	0	0	0
	7	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-1	0	0	0	-.5	0	-.5	0	0	0
		2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	.5	0	.5	0	0	0
	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-1	0	0	0	.5	0	.5	0	0	0
		2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	-.5	0	-.5	0	0	0
05	9	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0	-.5	0	0	-.5	0	0
		2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	.5	0	0	.5	0	0
	10	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0	-.5	0	0	-.5	0	0
		2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	.5	0	0	.5	0	0
	11	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0	-.5	0	0	-.5	0	0
		2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	.5	0	0	.5	0	0
	12	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0	.5	0	0	.5	0	0
		2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	-.5	0	0	-.5	0	0
08	13	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	0	-.5	0	0	0	-.5	0
		2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	.5	0	0	0	.5	0
	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	0	0	0	-1	0	-.5	0	0	0	-.5	0
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	.5	0	0	0	.5	0
	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	-1	0	-.5	0	0	0	-.5	0
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	.5	0	0	0	.5	0
	16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	-1	0	0	0	-1	0	.5	0	0	0	.5	0
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	-.5	0	0	0	-.5	0
11	17	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	-1	0	0	0	0	-1	-.5	0	0	0	0	-.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	.5	0	0	0	0	.5
	18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	-1	0	0	0	0	-1	-.5	0	0	0	0	-.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	.5	0	0	0	0	.5
	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	-1	0	0	0	0	-1	.5	0	0	0	0	.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	-.5	0	0	0	0	-.5
	20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-1	0	0	0	0	-1	.5	0	0	0	0	.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	-.5	0	0	0	0	-.5
13	21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	1	1	1	1	1	-.5	.5	.5	.5	.5	.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
	22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	-1	1	1	1	1	1	-.5	.5	.5	.5	.5	.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	-1	-1	-1	-1	-1	.5	-.5	-.5	-.5	-.5	-.5
	23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	1	1	1	1	1	.5	-.5	-.5	-.5	-.5	-.5
		2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5
	24	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	.5	-.5	-.5	-.5	-.5	-.5
		2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-.5	.5	.5	.5	.5	.5

Table VII - 18

Statistics from the Regression Analyses of Spring WRAT Scores
for the Multiple Cohort Study

Regression Statistics	Hypothesis Tested	F-Ratio. Refer Roman numerals to first column	F	df	p
Analysis A	Sponsor Effects	$\frac{(III - IV)/5}{(IX - I)/11}$	3.057	5,11	0.05
I. $R^2 y.G,H,I,HxI$ = .7789					
II. $R^2 y.G,H,I$ = .7534	FT/NFT Effects	$\frac{(II - III)/1}{(IX - I)/11}$	.330	1,11	
III. $R^2 y.G,H$ = .7505	Sponsor x FT/NFT Effects	$\frac{(I - II)/5}{(IX - I)/11}$	.580	5,11	
IV. $R^2 y.G$ = .6160					
Analysis B					
V. $R^2 y.G,J,K,HxK,$ IxK,HxIxK = .9524	Cohort Effects	$\frac{(VIII - IX)/1}{(1.- V)/12}$	1.100	1,12	
VI. $R^2 y.G,J,K,HxK,$ IxK = .9273	Sponsor x Cohort Effects	$\frac{(VII - VIII)/5}{(1.- V)/12}$	2.290	5,12	0.10
VII. $R^2 y.G,J,K,$ HxK = .9262	FT/NFT x Cohort Effects	$\frac{(VI - VII)/1}{(1.- V)/12}$	.275	1,12	
VIII. $R^2 y.G,J,K$ = .8804	Sponsor x FT/NFT x Cohort Effects	$\frac{(V - VI)/5}{(1.- V)/12}$	1.255	5,12	0.25
IX. $R^2 y.G,J$ = .8760					

5.7.0 RESULTS

The results of analysis A, as displayed in this table, indicate that Sponsors differ in their effects on the covariate adjusted scores regardless of the cohort membership. Analysis B informs us that there is a barely significant (.25 level) triple interaction, Sponsor by FT/NFT by cohort, suggesting that the Sponsor by treatment effect differs across cohorts. In the light of this weak relationship, it is reasonable to turn to the two-way interactions. Here, the Sponsor by cohort effect is significant at the .10 level. Sponsorship produces different effects on Chort I and Cohort III.

The data indicate that the Fall to Spring adjusted slope is somewhat steeper for Cohort III than for Cohort I despite the generally lower Fall scores for Cohort III. Figure VII-44 indicates that this trend is found more often in the FT groups than the NFT groups which accounts for the three way, Sponsor x FT/NFT x cohort interaction. The significant Sponsor x cohort interaction may be interpreted as an indication that Sponsors were more effective with Cohort III than Cohort I on the WRAT. This supports the notion that early attempts to implement programs involved problems which may have interfered with some aspects of the models.

5.8.0 DISCUSSION

This interpretation is offered with a great deal of tentativeness since a variety of other factors may be operating which distinguish between the events occurring in 1969-70 and those occurring in 1971-72. There are children, teacher, and community differences which have not been accounted for. It might also be true that real problems of implementation might not emerge for some Sponsors until after several years of experience with their models in the field. These issues need to be examined in greater detail before we fully accept the notion that Sponsors get better (in respect to scores on the WRAT) over time. It is true, however, that at this point in the longitudinal study, it is certainly not appropriate to reject the hypothesis that Sponsor maturation is positively related to WRAT scores.

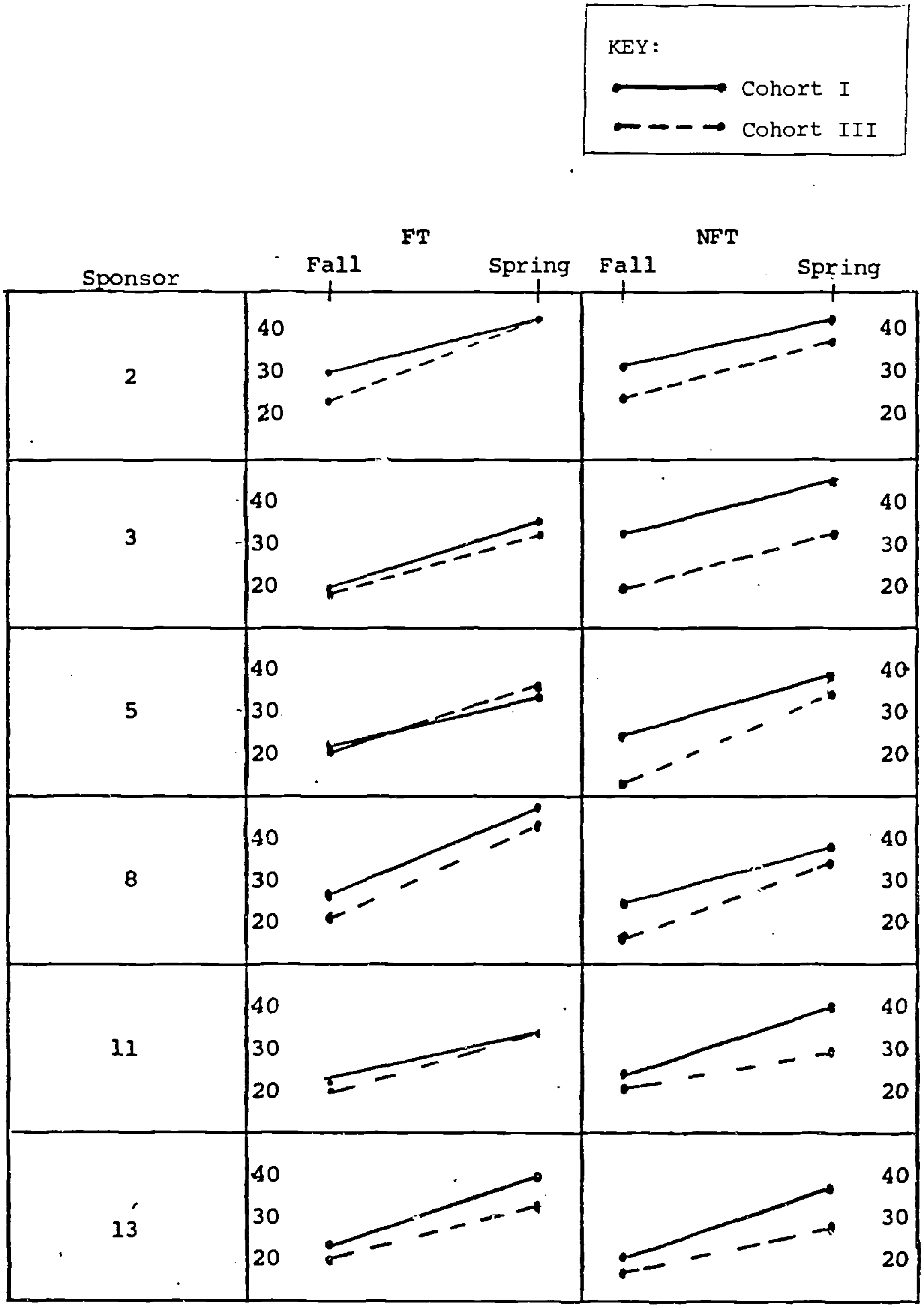


Figure VII - 44. Comparison of Cohort I and Cohort III One Year Kindergarten Gains on WRAT Scores for Follow Through and Non-Follow Through Schools in Six Sponsors

One factor that can be examined further in this matter is the testing schedules for both cohorts. Recent examintions of new data indicate that the variable "pretest delay" which is the number of days from the beginning of a given school to the time of pretest administration, varies systematically across these two cohorts. Cohort I was tested considerably later in the year than Cohort III. Thus, the fact that Cohort I had higher scores than Cohort III on the Fall WRAT may reflect the testing schedules rather than a difference in true pretest levels of the two groups. Although the pretest scores are used as a covariate, the full information relating to the difference between the Fall performance of the two cohorts may not be fully represented therein. Consequently, analysis B was replicated with the pretest delay included as an additional covariate in set G. Table VII-19 shows the results.

The inclusion of the testing schedule into the model does not alter the significance of the Sponsor x FT/NFT x cohort interaction. Turning to the two-way interaction we find that the significance level of the Sponsor by cohort effect is reduced from .10 to .25 when we control for the different testing schedules. The message here is that Cohort III may have received a small unwarranted advantage when the Fall WRAT scores were used to adjust for initial differences. Compensating for this possible error by adjusting for pretest delay serves to slightly reduce the differential Sponsor effects across cohorts. This reduction in effects might be accounted for by systematic differences in testing schedules across the two cohorts.

Two facts prohibit us from rejecting the hypothesis of Sponsor maturation as studied here: (1) Both the two-way and three-way interactions are issuing slight signals amidst the loud noise apparent in this model; (2) We are uncertain as to how the pretest delay is operating in the model. The question of the effect of Sponsor maturation is an important issue which at this point in time must remain among the viable hypotheses in need of further examination.

Table VII - 19

Statistics from Regression Analysis B
for the Multiple Cohort Study
with Pretest Delay Included
in Set G

Regression Statistics	Hypothesis Tested	F-Ratio: Refer Roman numerals to first column	F	df	p
V. $R^2 y.G,J,K,HxK, IxK,HxIxK = .9536$	Cohort Effects	$\frac{(VIII - IX)/1}{(1. - V)/12}$	.410	1,12	
VI. $R^2 y.G,J,K,HxK, IxK = .9276$	Sponsor x Cohort Effects	$\frac{(VII - VIII)/5}{(1. - V)/12}$	1.590	5,12	0.25
VII. $R^2 y.G,J,K,HxK = .9262$					
VIII. $R^2 y.G,J,K = .8952$	FT/NFT x Cohort Effects	$\frac{(VI - VII)/1}{(1. - V)/12}$	.359	1,12	
IX. $R^2 y.G,J = .8760$	Sponsor x FT/NFT x Cohort Effects	$\frac{(V - VI)/5}{(1. - V)/12}$	1.333	5,12	0.25

6.0 CONCLUDING STATEMENTS

The search for FT/NFT contrasts at the school level of aggregation has, at this point in the evaluation effort, raised as many issues as it has appeared to resolve. Every contrast for every Sponsor clearly needs to be interpreted in the light of local site conditions. The large variation in FT/NFT mismatch across Sponsors suggests that the criteria for inclusion in the FT or NFT groups varied a great deal from one Sponsor to the next, which produced more than statistical artifacts. The central conditions under which the Sponsors implemented their programs varied, so that the meaning of the contrasts changed from one Sponsor to the next. The fact that some Sponsors (e.g., University of Oregon) demonstrated consistently higher adjusted achievement scores across a variety of sites, while some (e.g., Bank Street) showed no achievement gains relative to the local NFT groups, must be examined in the light of program as well as model factors. The Bank Street sites were relatively more affluent than those met by several Sponsors, and their FT children were drastically lower on achievement scores when they entered FT than the FT children assigned to many other Sponsors. There were reasons for these differences (as yet unknown to the present writers) which must have influenced the way in which the Bank Street personnel had to deal with those sites. Such programmatic factors must be different than those found at the sites of other Sponsors. The contrasts between FT and NFT schools cannot be fully understood without knowledge of these factors.

The rather diminished FT contrasts found in the Big Cities further support this notion of the importance of site-specific factors. Innovative programs, covering the wide range of approaches represented by the six Sponsors operating in New York, Philadelphia, and Chicago, were all less effective generally than they were outside these cities. In some cases the pattern of effects was sharply changed in the Big Cities, compared to other sites. This is not likely to be a random effect, and it may not be fully attributable to the staff operating the programs in the Big Cities. It is just as likely attributed to the nature of the children in these cities, the nature and expectations of their parents, the structure of the school systems, the nature of school-community

relations, and the nature of the programs available to the NFT schools in the Big Cities. These are factors which go well beyond the Sponsor models as explanatory factors, and which must be explored carefully in order to understand the impact which innovative programs have upon school systems.

Still another factor which highlights the site-specific issues is the time of testing data reported here. Although this study is incomplete because many of the required data are not yet available, it is clear that at some sites there was a strong tendency for higher scoring schools to be tested later in the Spring testing period than lower achieving schools. In addition, there is some indication that for a few Sponsors there is a relationship between how far into the school year the pretest was administered and the magnitude of that pretest score. This latter point is not likely to be accounted for by early treatment effects, although it might reflect the adaptation of children to the school situation, which might in turn contribute to test performance. This is not likely to account for some of the negative relationships observed, so that at least one further hypothesis remains to be seriously considered. This has to do with the local conditions which contributed to the testing schedule. A testing schedule in which higher achieving schools tested earlier at some sites and later at other sites reflects some as yet unknown but rather subtle school and community factors impinging on the performance of children.

Despite these caveats, it is clear that both achievement and affective effects attributable to different FT Sponsors are to be found in these data. FT kindergarten children do appear to be engaged in experiences which are meaningfully different than those of their NFT mates. In addition, the FT effect is greater overall in 1971-72 than in 1969-70, and this may be attributable to the experiences Sponsors have had over the years both in implementing their programs and designing their models to fit the needs of local conditions. There is no doubt that site-specific issues of program implementation must be added to these analyses in order to make more sense out of these data. At the same time, it is clear that Sponsors need to be examined in the light of the kinds

of classes and children with whom they are dealing before their impacts can begin to become apparent. Site conditions and child and classroom properties are all factors which must be studied with, as well as partialled from, Sponsor effects. Information on site conditions is not yet available for analysis, but a selected set of class and child variables are present in the data; it is the interactions of Sponsors with these factors to which we now turn.

CHAPTER VIII

CLASS AND CHILD VARIABLES AFFECTING FT/NFT CONTRASTS

The studies reported here represent the first set of approaches to the basic concerns of the national evaluation as these are conceived by the present writers, namely, subject by treatment interactions. These studies initiate the search for critical interactions utilizing variables of major interest. These include the entering achievement levels of the classes, the ethnic mix of the classes, the ethnic membership of the children, the sex of the children, and the preschool experience of the children. Future studies will also include an examination of kinds of children within kinds of classrooms interacting with Sponsors. As we reach that level of complex analyses, we shall be approaching the most informative areas of study for both theoretical and practical concerns. The present studies should be taken as the first steps in this direction.

1.0 CLASS ETHNIC COMPOSITION

1.1.0 INTRODUCTION

This study explores the relationship between the ethnic composition of a class and class performance on the outcome measure. The literature indicates that minority children, particularly Black children in integrated classes in upper elementary grades, perform higher on achievement measures than comparable children in segregated classes (Coleman et al., 1966; McPartland, 1968). Thus, the ethnic composition of the classroom is of interest not only as a potential correlate of an advantageous educational situation for minority kindergarten children, but also as a potentially confounding factor in the Follow Through evaluation. The present study addresses two basic questions:

- Does the ethnic composition of the class relate to class performance?
- Do Sponsors have different effects on classes which are integrated to different degrees?

The Follow Through data, including data on a large number of kindergarten classes on both achievement and affective outcomes, provide a good

opportunity to study the general effect of the ethnic composition of a class on its performance. Previous studies suggest that integration confers few achievement advantages on classes before the third grade, and also that fate control or locus of control is more highly correlated with achievement in Black children than in White children (Coleman et al., 1966).

A number of theories have been developed that relate these two domains and generally suggest that the acquisition of a sense of control of one's fate is a developmental process that results from the internalization of value systems and from the formation of expectancies derived from specific experiences. One's performance on achievement tasks is then determined partially by one's ability, partially by one's ability self-concept, and partially by one's self-efficiency (Katz, 1968).

These theories suggest that the attitudes and behaviors displayed in integrated classes present an environment that is appropriate for the academic growth of minority children. This study does not test any of the complex hypotheses that have been developed in this area. It simply asks the question: what does Follow Through do to mean achievement and affective levels in kindergarten classes of different ethnic compositions?

The second question this study addresses concerns the potential confounding of the Sponsors' effects and classroom ethnic composition effects. If we find that classes with a mix of majority and minority children have higher average scores on some measures and that their classes are not distributed uniformly across Sponsors, then the positive effects of Sponsors with such mixed classes must be attributed in part to the heterogeneity of distribution of such classes across Sponsors. The present study is designed to explore the presence of such confounding.

1.2.0 METHOD

1.2.1 Analytic Subset

A total of 404 classes distributed across Sponsors' FT and NFT groups were used in this study. The distribution of classes on ethnic composition within Sponsor is shown in Table VIII-2 and will be discussed below. Each class in these analyses contains at least five children, all of whom had

complete data on all outcome and background variables. Greater detail on the composition of the sample is presented in Chapter III.

1.2.2 Measures

The covariables included the class mean Fall WRAT; the class means for mother's education, years at address, adjusted income index, city size, teacher's ethnicity, teacher's education, teacher's experience, parent's perceptions of the receptivity of the school, and parent participation. A detailed description of these covariables is found in the section on covariables. The outcomes included Spring class means for the WRAT; the MAT Reading, Arithmetic, and Listening to Sounds subtests; the Gumpgookies; the Locus of Control means; and Absence.

1.2.3 Analytic Method

The multiple regression analogue of ANCOVA was used to estimate regression coefficients and variance components for two hypotheses. The first, a linear hypothesis, suggests that there is a uniform change in an outcome as the proportion of white children in a class increases. The confirmation of this hypothesis (finding that a significant proportion of variance is accounted for by a linear fit of proportion white in class on an outcome) suggests that there is a component of an outcome that relates directly to the number of white children in a class. In addition, any departure from a slope of zero indicates a differential effect as a function of class ethnic composition and the possibility of confounding within Sponsors. The second hypothesis, the nonlinear hypothesis, suggests that the change in an outcome as proportion of white children in class increases is not uniform and that classes with a mix of majority and minority students perform differently from predominantly nonwhite classes and perhaps differently from predominantly white classes. This brief statement does not exhaust the possible interpretation of a curvilinear fit on the data, but does follow from the previous findings in the literature. Confirmation of this hypothesis would suggest that classroom racial composition is related to an outcome and that Sponsors' effects may be confounded with this effect.

The assessment of the linear fit of proportion white in a class to an outcome needs little comment since the procedures followed are identical to the assessment of any graduated variable. The variable, proportion of

white children in a class, an effects coded variable representing FT/NFT membership, and a set of nine effects coded variables representing Sponsor membership are used to partition the variance in an outcome in an order indicated by the hypothesis. Since the effects of ethnic composition in a class are of interest, the variance accounted for by the covariable, and other main effects (FT/NFT membership and Sponsor membership), are partitioned prior to the assessment of the variance attributable to ethnic composition. This hierarchical order yields a unique variance component, and a semi-partial correlation for ethnic composition. The significance of the variance component is then assessed relative to the error variance using a conventional F test. For the nonlinear hypothesis, the procedure is identical.

The nonlinearity is represented by a single term: the square root of the proportion of white in class. Under this nonlinear hypothesis, the change in Y with a change in X is smaller as X increases, corresponding to the idea that classes with mixed ethnic composition are more like predominantly white classes than predominantly nonwhite classes.

The assessment of the appropriateness of the fit of this transformation is accomplished in the same manner as for the linear fit. The variable representing the nonlinear fit is entered into a regression equation after other relevant factors have already been entered and the increment in explained variance is assessed using a conventional F test. Since the present study is concerned with the question of whether a linear or nonlinear hypothesis is appropriate, and since the nonlinear fit could have a linear component, the nonlinear factor is entered into the predictive equation after the linear factors. The hierarchical model thus includes a set of covariables, an effects coded variable representing FT/NFT membership, a variable representing proportion white in a class (linear component), and a variable representing the square root of proportion white in a class (nonlinear component). These factors are entered into the regression equation in the order indicated above. Interactions of these factors, due to the generally small number of classes within Sponsor, were not included in the model (see below). The model, the hierarchical order of variance partitioning, and the F ratios utilized are shown in Table VIII-1. In the results section, both of these hypotheses are explored and within-Sponsor effects are considered.

TABLE VIII - 1

F Ratios and Factors in the Hierarchical Model of ANCOVA Ethnic Composition Study

F RATIOS: Linear

$$F = \frac{sr_A^2 \div 1}{(1 - R^2_{Y \cdot MAIN}) \div 381}$$

Nonlinear

$$F = \frac{sr_A^2 + sr_B^2 \div 2}{(1 - R^2_{Y \cdot MAIN}) \div 380}$$

ANCOVA FACTORS:

Covariates = Fall Wrat
Years at Address
Adjusted Income Index
City Size
Teacher's Education
Teacher's Experience
Teacher's Ethnicity
Level of Parent Participation
Parent Perceived Receptivity of School

A Proportion White in Class
B Square Root of Proportion of White in Class
C Sponsor = 2,3,5,7,8,9,10,11,12,14
D FT/NFT
E Predictor by Sponsor
F Predictor by FT/N FT
G Sponsor by FT/NFT
H Predictor by Sponsor by FT/NFT

$$sr_A^2 = R^2_{Y \cdot cov\ ACD} - R^2_{Y \cdot cov\ CD}$$

$$R^2_{Y \cdot MAIN} = R^2_{Y \cdot cov\ ABCD}$$

$$sr_B^2 = R^2_{Y \cdot cov\ ABCD} - R^2_{Y \cdot cov\ ACD}$$

sr^2 represents the squared semi-partial correlation or the percent of the variance uniquely accounted for by the factor indicated.

1.3.0 RESULTS

The distribution of classes across five categories of ethnic composition is presented in Table VIII-2 for pooled FT and NFT groups and for each Sponsor's FT and NFT groups. The table indicates that the pooled FT and NFT groups are fairly comparable in their distribution of classes across categories of proportion white in class. However, the FT group has approximately three times as many classes as the NFT group in the predominantly nonwhite category and this ratio of 3 to 1 is repeated in many of the individual Sponsors' distributions.

Although the overall distribution of classes is bimodal with the highest number of classes falling in the predominantly white or predominantly nonwhite categories, there is a sufficiently large number of classes in the mixed categories to permit an overall analysis. The proportions of variance accounted for by the linear factor, and by the linear plus the non-nonlinear factors, are presented in Table VIII-3 along with the total proportions of variance accounted for by main effects, and the corresponding F ratios and significance levels, for all eight outcome measures.

The linear hypothesis that outcomes change uniformly with changes in the ethnic composition of the class is supported only for the Locus of Control for positive events. The regression coefficient for proportion white in class and class Locus of Control for positive events was .17, indicating a weak positive mean relationship: classes with a higher concentration of white pupils also exhibit higher scores (more internal) on positive Locus of Control.

The nonlinear hypothesis fits the data as badly as the linear hypothesis for the achievement outcomes. However, for two of the affective measures and the Absence outcome, the nonlinear hypothesis is supported. The obtained relationships are shown in Figures VIII-1a through 1c. Figure VIII-1a shows adjusted class mean Locus of Control for positive events as a function of percent white in class (least-squares fit). The effects of the covariables, FT/NFT membership and Sponsor membership have already been partialled out of this relationship.

TABLE VIII - 2

Number and Percent of Total Numbers of Classes Falling Into Each of Five Categories of Percent White in Class for Each Sponsor's FT and NFT Group and for the Pooled FT and NFT Groups.

Sponsor	< 20%	> 20% < 40%	> 40% < 60%	> 60% < 80%	> 80%
2 FT	18	16	4	4	5
NFT	11	1	3	1	13
3 FT	13	2	5	9	9
NFT	4	3	1	5	13
5 FT	8	0	1	4	12
NFT	0	0	2	1	8
7 FT	15	2	1	2	1
NFT	7	1	0	2	2
8 FT	30	0	1	3	0
NFT	9	0	1	2	0
9 FT	18	6	1	0	0
NFT	6	6	0	1	4
10 FT	14	0	2	2	5
NFT	3	0	2	1	3
11 FT	12	1	0	0	7
NFT	7	0	1	0	7
12 FT	6	0	0	0	14
NFT	1	2	0	0	8
14 FT	6	1	5	0	0
NFT	0	3	3	1	0
TOTAL FT	126	28	15	33	53
NFT	48	16	13	13	58
TOTAL	174	44	28	46	111

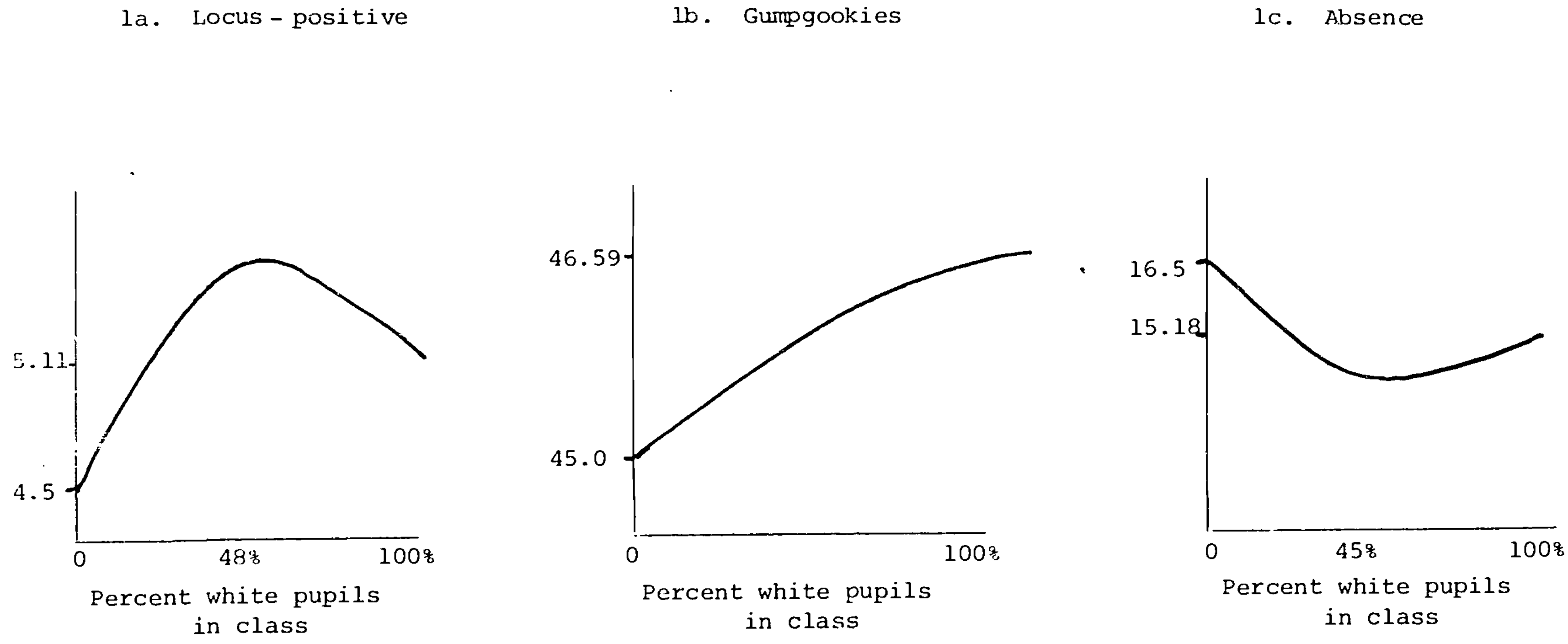
TABLE VIII - 3

Summary of Effects, Class Level Analysis

Proportion of White in Class

Proportion of White in Class + Squared Root of White in Class

	WRAT	MAT			GUMP-GOOKIES	LOCUS OF CONTROL		ABSENCE
		LISTENING	READING	MATH		Positive	Negative	
Proportion White in Class d.f. = 1,381								
sr_G^2	0.001	0.003	0.001	0.000	0.006	0.018	0.004	0.000
$R^2_{Y \cdot MAIN}$	0.652	0.427	0.446	0.479	0.212	0.302	0.116	0.208
F	0.855	2.267	0.756	0.022	2.926	9.818	1.763	0.188
p	NS	NS	NS	NS	0.100	0.005	NS	NS
Proportion White in Class + Squared Root of Proportion of White in Class d.f. = 2,380								
sr_G^2	0.003	0.004	0.002	0.003	0.015	0.040	0.001	0.018
$R^2_{Y \cdot MAIN}$	0.607	0.392	0.474	0.379	0.187	0.315	0.118	0.192
F	1.406	1.121	0.836	0.855	3.401	11.058	2.128	4.275
p	NS	NS	NS	NS	0.050	0.005	NS	0.025



FIGURES VIII - 1a, 1b, and 1c

Main Effects of Classroom Ethnic Composition Across All Sponsors for Locus-positive, Gumpgookies and Absence.

Cohort III, Kindergarten, Class level N = 404

The curve suggests that integrated classes feel more personal control of the good things that happen to them than do predominantly Black classes, and at least as much so as do predominantly white classes. The linear hypothesis is also represented in this figure. The predominantly White classes have higher scores than the predominantly Black classes, but the major portion of variance is contributed by the nonlinear aspect.

Figure VIII-1b shows the relationship between class mean Gumpgookies and proportion of White in the class. The relationship is similar to that found with the locus measure. As proportion White in class increases, achievement motivation increases but the rate of increase decreases. That is, the integrated classes are more like the predominantly White classes than the predominantly black classes.

Finally, Figure VIII-1c shows the relationship between class mean Absence and percent white in class. The curve indicates an opposite algebraic relationship from that found with the other measures. The predominantly nonwhite classes have the highest mean Absence score, the integrated classes have the lowest mean, and the predominantly White classes have an intermediate value somewhat closer to that of the integrated group than to that of the predominantly nonwhite group.

The integrated classes have children with stronger academic motivation, more internalization of responsibility, and fewer absences than either predominantly white or predominantly nonwhite classes. This is clearly a function of the mix in the classroom rather than the unique properties of either White or Black groups; that is to say, the nonlinear hypothesis was more strongly supported than the linear. Social and community factors which may have contributed to these scores, and which may be different in locales where integration takes place than in locales where integration does not occur, were probably not completely accounted for by the covariable set used in the present study. We conclude, therefore, that in communities where integrated kindergartens are to be found, motivational advantages are present which must be carefully observed through subsequent grades.

We now turn to the assessment of within-Sponsor class ethnicity effects. For most Sponsors, the ethnicity distribution of classes is far too skewed to permit a meaningful analysis. Sponsors 5, 10, 11, and 12 have fewer than a quarter of their classes in the middle ethnicity range (proportion White > 20% and ≤ 80%). Sponsors 7, 8, 9, and 14 have a very small percentage of their classes in the predominantly White (> 80%) category. Both the small numbers of observations within these Sponsors' groups as well as their skewed ethnicity distributions obviate the assessment of ethnic composition effects within these Sponsors.

The remaining Sponsors, 2 and 3, have predominantly nonwhite classes in their FT groups with their other classes spread more or less evenly across the range of proportion White in class. Although these Sponsors have some spread in their distribution of FT classes across proportion White in class, the number of classes in each category is quite small. We tested, however, the fit of the linear hypothesis for these groups, using a within-Sponsor design identical to the main effect design. The analyses indicate that the linear fit of outcomes on percent White in class did not vary from the overall result. That is, the overall rejection of the linear hypothesis was not altered by the consideration of the within-Sponsor effects.

The nonlinear hypothesis was also assessed for these Sponsors' FT groups and again the general rejection of the hypothesis was confirmed for the achievement measures and Locus of Control for negative events. On the three outcomes for which the nonlinear hypothesis was supported as a main effect, there were no within-Sponsor differences for these two Sponsors. Thus, the overall picture was not altered by looking at effects within Sponsors.

1.4.0 DISCUSSION

In general, the results suggest that integration may produce affective advantages as well as a reduced absence rate. The results for Gumpgookies, Locus of Control for positive events, and Absence all indicate favorable performance in integrated kindergarten classes. This affective development could be a key to future achievement and may represent a substantial

disruption of the basis of the traditional academic decline in minority children. We, however, are aware that the low reliability of those affective measures may be contributing some interesting but spurious results.

At best these results and this study must be looked at as exploratory on the one hand and inconclusive on the other. The sample does not generally permit the assessment of effect by Sponsor; therefore, biases may be present for which appropriate adjustments cannot be made. Further, the assessment of the overall integration effect is tenuous because of the pooling of potentially confounding factors and the disproportionality of the distribution. However, the effects on the Gumpgookies, Locus of Control, and Absence outcomes are anticipated by previous theoretical work and the results provide a promising background for future research.

Ethnic composition of class is a variable which accounts for some class variance in domains of major interest to all Sponsors. It is extremely unfortunate that the sample does not permit assessment of effect by Sponsor. This is particularly true because the only two Sponsors with enough integrated classes to participate in the examination of the interaction term, University of Arizona and Far West Laboratory, are very similar in their curriculum approaches. It would have been quite instructive, both for an understanding of the dynamics of integrated classes, and for the assessment of Sponsor effects, to examine a variety of Sponsors interacting with this variable. It is hoped that in the future, more integrated classes associated with more Sponsors will be found in the data base. In the meantime, it is clear that ethnically mixed classes might provide Sponsors with potentially fertile ground upon which to cast their innovative seed.

2.0 ENTRY LEVEL STUDY

2.1.0 INTRODUCTION

The entry level study is concerned with the relationship between a class's initial achievement level (mean Fall WRAT) and posttest performance level within a Sponsor's FT and NFT classes. All other studies in this report utilize the Fall WRAT as the primary covariable. The present study is very different in this respect, in that the relationship between a class's initial achievement level and outcomes is explored. Specifically, this study addresses the questions:

- Are the relationships between the initial achievement level and posttest performance in a Sponsor's FT and NFT classes sufficiently similar to justify adjustment for initial difference and comparison?
- Do some Sponsors have systematically different effects on classes that start out at different levels of achievement?

The former question addresses the issue of homogeneity of regression of an outcome on initial achievement. Homogeneity of regression (that is, a uniform relationship between an outcome and initial achievement within a Sponsor's FT and NFT classes) is a prerequisite for the use of initial achievement level as a covariable and for the adjustment of initial differences on achievement. The absence of such uniformity (heterogeneity of regression) restricts the exploration of the Sponsor's effectiveness in that we cannot assess the difference between FT and NFT classes independent of the initial achievement differences. The results presented in this section indicate that the class effects of several Sponsors are confounded by initial achievement differences. These results only apply to the class level studies and do not relate directly to the school or child studies. This does not obviate the exploration of Sponsor's effect, but complicates the exploration of the initial achievement level of a class and must be considered in the exploration of gain. We can accomplish this by exploring the relationship between initial achievement and a covariable-adjusted posttest score, that is, the amount of gain[1] that a class shows can be explored relative to the

[1]Gain here refers to a posttest score adjusted by all covariables except the Fall WRAT. This is the essential differences between this and other studies in this report. All other covariables were found to be homogeneous across Sponsors FT and NFT groups.

initial achievement level. This brings us to the second question. Given the finding that within a Sponsor's FT or NFT classes the amount of gain that a class shows is dependent on its entry level, then we must assess what kind of classes are showing what kind of advantages or disadvantages.

2.2.0 METHOD

2.2.1 Analytic Subset

A total of 404 FT and NFT classes from Cohort III kindergarten were used in this study. Each class represents an aggregate of no fewer than five children, all of whom had complete information on all of the background and outcome variables.

2.2.2 Measures

The covariables included the class means for mother's education, years at address, parent perceived receptivity of the school, and parent participation; adjusted income index; city size; teacher's ethnicity; teacher's education; teacher's experience; and percent white in class. A detailed description of these covariables is found in the section on covariables. The outcomes included class means for the Spring WRAT; the MAT reading, arithmetic, and listening to sounds subtests; the Gumpgookies; the Locus of Control means; and Absence.

2.2.3 Analytic Method

The multiple regression formats of ANOVA and ANCOVA were used to estimate unadjusted and covariable-adjusted effects, respectively. The model included a single graduated predictor representing initial achievement level (class mean, Fall WRAT); a single effects coded predictor representing FT/NFT membership; a set of nine effects coded predictors representing Sponsor membership; and the various two- and three-way interactions of these sets. The model, the hierarchical order of variance partitioning, and F ratios, are shown in Table VIII-4. The terms of interest in the present study are specified in the three-way interaction set, initial achievement level by FT/NFT membership by Sponsor membership.

TABLE VIII - 4

F Ratios and Factors in the Hierarchical Model of ANOVA and ANCOVA

Entry Level Study

F RATIOS: $$F = \frac{sr_G^2 \div 9}{(1 - R^2_{Y \cdot TOTAL}) \div 364}$$

ANOVA FACTORS:

A Predictor = Initial Achievement Level
B Sponsor = 2,3,5,7,8,9,10,11,12,14
C FT/NFT
D Predictor by Sponsor
E Predictor by FT/NFT
F Sponsor by FT/NFT
G Predictor by Sponsor by FT/NFT

$$sr_G^2 = R^2_{Y \cdot ABCDEFG} - R^2_{Y \cdot ABCDEF}$$

$$R^2_{Y \cdot TOTAL} = R^2_{Y \cdot ABCDEFG}$$

F RATIOS: $$F = \frac{sr_G^2 \div 9}{(1 - R^2_{Y \cdot TOTAL}) \div 353}$$

ANCOVA FACTORS:

Covariates = Mother's Education
Years at Address
Adjusted Income Index
City Size
Teacher's Education
Teacher's Experience
Teacher's Ethnicity
Percent White in Class
Level of Parent Participation
Parent Perceived Receptivity of School

A Predictor = Initial Achievement Level
B Sponsor = 2,3,5,7,8,9,10,11,12,14
C FT/NFT
D Predictor by Sponsor
E Predictor by FT/NFT
F Sponsor by FT/NFT
G Predictor by Sponsor by FT/NFT

$$sr_G^2 = R^2_{Y \cdot cov\ ABCDEFG} - R^2_{Y \cdot cov\ ABCDEF}$$

$$R^2_{Y \cdot TOTAL} = R^2_{Y \cdot cov\ ABCDEFG}$$

sr^2 represents the squared semi-partial correlation or the percent of the variance uniquely accounted for by the factor indicated.

2.3.0 RESULTS

The amount of variance accounted for by the set of three-way interaction terms, the total variance accounted for, an F ratio, and its significance are presented in Table VIII-5 for each outcome for the ANOVA and ANCOVA models.

In general, the relationships between initial class achievement level and the outcomes are uniform. The Spring WRAT is the only exception to this in the ANOVA model, while in the ANCOVA model all of the achievement outcomes with the exception of the MAT arithmetic outcome indicate some non-uniform regression between some Sponsors' FT and NFT classes. In order to present these results as meaningfully as possible, the effects for each Sponsor for each outcome on which non-uniform regression occurs are presented individually along with relevant sampling information, including the distribution of classes across five categories of initial achievement. The five categories are defined by intervals in total sample standard deviation units around the total sample mean on initial achievement level (S.D. = 6.19; Mean = 35.58). Although this categorization is not totally appropriate, since initial achievement level is utilized as a continuous variable in the analyses, the categorization permits an appreciation of the range and distribution of the classes on initial achievement as well as an exploration of the appropriateness of the regression estimate obtained from the analysis.

2.3.1 Sponsor 2: Far West Laboratory

The distribution of Sponsor 2's FT and NFT classes across five categories of initial achievement, and the overall FT and NFT means and standard deviations of initial achievement level, are shown in Tables VIII-6 and VIII-7. The tables indicate that the two groups of classes are quite comparable on their overall means and standard deviations, as well as on their initial achievement levels.

The ANOVA model yields a regression coefficient of 1.08 for the FT and .71 for the NFT groups. These values are substantially different ($F = 6.502$; $df = 1, 346$; $P < .025$) and indicate that the use of the Fall WRAT as

TABLE VIII - 5

Class Entry-Level Study: Variance Accounted for by the Interaction of Initial Achievement Level, Sponsor Membership, FT/NFT Membership; Total Variance Accounted for, F Ratio, and Significance for the ANOVA and ANCOVA Models.

	WRAT	MAT			GUMP-GOOKIES	LOCUS OF CONTROL		ABSENCE
		LISTENING	READING	MATH		Positive	Negative	
ANOVA	Entry-Level by Sponsor by FT/NFT Interaction (d.f. = 9,364)							
sr^2_G	0.031	0.025	0.015	0.023	0.022	0.015	0.019	0.016
$R^2_{Y \cdot TOTAL}$	0.698	0.505	0.538	0.464	0.183	0.304	0.158	0.189
F	4.186	1.844	1.339	1.703	1.068	0.849	0.918	0.785
p	0.005	NS	NS	NS	NS	NS	NS	NS
ANCOVA	Entry-Level by Sponsor by FT/NFT Interaction (d.f. = 9,353)							
sr^2_G	0.030	0.024	0.011	0.023	0.019	0.018	0.015	0.021
$R^2_{Y \cdot TOTAL}$	0.718	0.524	0.581	0.503	0.261	0.368	0.179	0.266
F	4.118	1.969	1.032	1.847	1.004	1.130	0.730	1.122
p	0.005	0.050	NS	0.100	NS	NS	NS	NS

TABLE VIII - 6

The Distribution of Sponsors' FT/NFT Classes Across Five Categories of Initial Achievement Defined in Terms of the Sample Mean (35.68) and Standard Deviation (6.19) of Fall WRAT

Sponsor	FT/ NFT	$\leq \bar{X}$-2 SDs	$\geq \bar{X}$-2 SDs $< \bar{X}$-1 SD	$\geq \bar{X}$-1 SD $< \bar{X}$+1 SD	$\geq \bar{X}$+1 SD $< \bar{X}$+2 SDs	$\geq \bar{X}$+2 SDs
2	FT	0	6	25	5	1
	NFT	0	3	23	3	0
3	FT	1	3	29	3	2
	NFT	1	0	17	6	2
5	FT	1	5	0	17	2
	NFT	0	1	8	1	1
7	FT	0	1	14	6	0
	NFT	0	2	7	1	2
8	FT	1	2	27	3	1
	NFT	2	2	8	0	0
9	FT	1	3	20	1	0
	NFT	0	3	13	1	0
10	FT	0	5	15	2	1
	NFT	1	2	6	0	0
11	FT	0	3	17	0	0
	NFT	0	2	8	5	0
12	FT	0	2	11	3	4
	NFT	0	1	7	3	0
14	FT	1	2	9	0	0
	NFT	0	3	4	0	0

TABLE VIII - 7

Fall WRAT Means, Standard Deviation's and Number of Classes For Each Sponsors' Follow Through/non-Follow Through Groups

		SPONSOR										
		2	3	5	7	8	9	10	11	12	14	TOTAL
	N	37	38	25	21	34	25	23	20	20	12	255
FT	X	35.80	35.89	35.38	37.99	35.82	35.00	34.30	34.37	37.93	31.30	
	SD	5.96	6.26	6.12	5.15	5.72	5.34	5.90	4.92	7.84	4.95	
	N	29	26	11	12	12	17	9	15	11	7	149
NFT	X	36.10	38.59	37.74	36.06	30.48	35.04	32.00	36.35	36.52	29.43	
	SD	4.32	7.02	6.51	8.29	5.11	6.28	6.06	6.96	6.62	2.01	

covariable is inappropriate for the Spring WRAT outcome.

The regression lines of Spring on Fall WRAT from the ANCOVA model are presented in Figure VIII-2 for the FT and NFT classes and a relationship very similar to the ANOVA model is portrayed. The FT class with lower-than-average initial achievement levels produced lower Spring WRAT means than the NFT class with comparable Fall WRAT means, while an opposite relationship obtained for classes with higher-than-average Fall WRAT means. The departure of these two regression lines from parallel is not statistically significant; however, the pattern is replicated in the regression of the covariable-adjusted MAT reading subtest on the Fall WRAT shown in Figure VIII-3. The regression indicates a similar relationship to that found with the Spring WRAT; FT classes with a higher initial achievement level benefit more than classes with a lower initial achievement level relative to the NFT classes. Again the regression lines are not significantly different from parallel. The essential problem with the use of the Fall WRAT as a covariable with these outcomes is that the initial achievement score carries important information regarding Sponsor 2's effectiveness. Sponsor 2's program is apparently more effective with classes of higher-than-average initial achievement and less effective with classes of lower initial achievement. This interaction indicates that at least for the WRAT and MAT reading outcomes the program appears to be ineffective overall, when in fact some types of classes may benefit from such a program.

2.3.2 Sponsor 3: University of Arizona

Sponsor 3's FT and NFT classes have a similar spread in their distribution across initial achievement level and similar standard deviation, but the NFT group has a 2.7 point average overall. (See Tables VIII-6 and VIII-7.) The NFT group also has a relatively higher percentage of its classes in the higher initial achievement level categories. The regression of the Spring WRAT on Fall WRAT for the FT group is .75 and for the NFT group .57. Although the difference between these slopes is large, the difference is not statistically significant ($F = 2.56$; $df = 1, 364$). Furthermore, the FT regression coefficient is similar to the overall coefficient (.74)

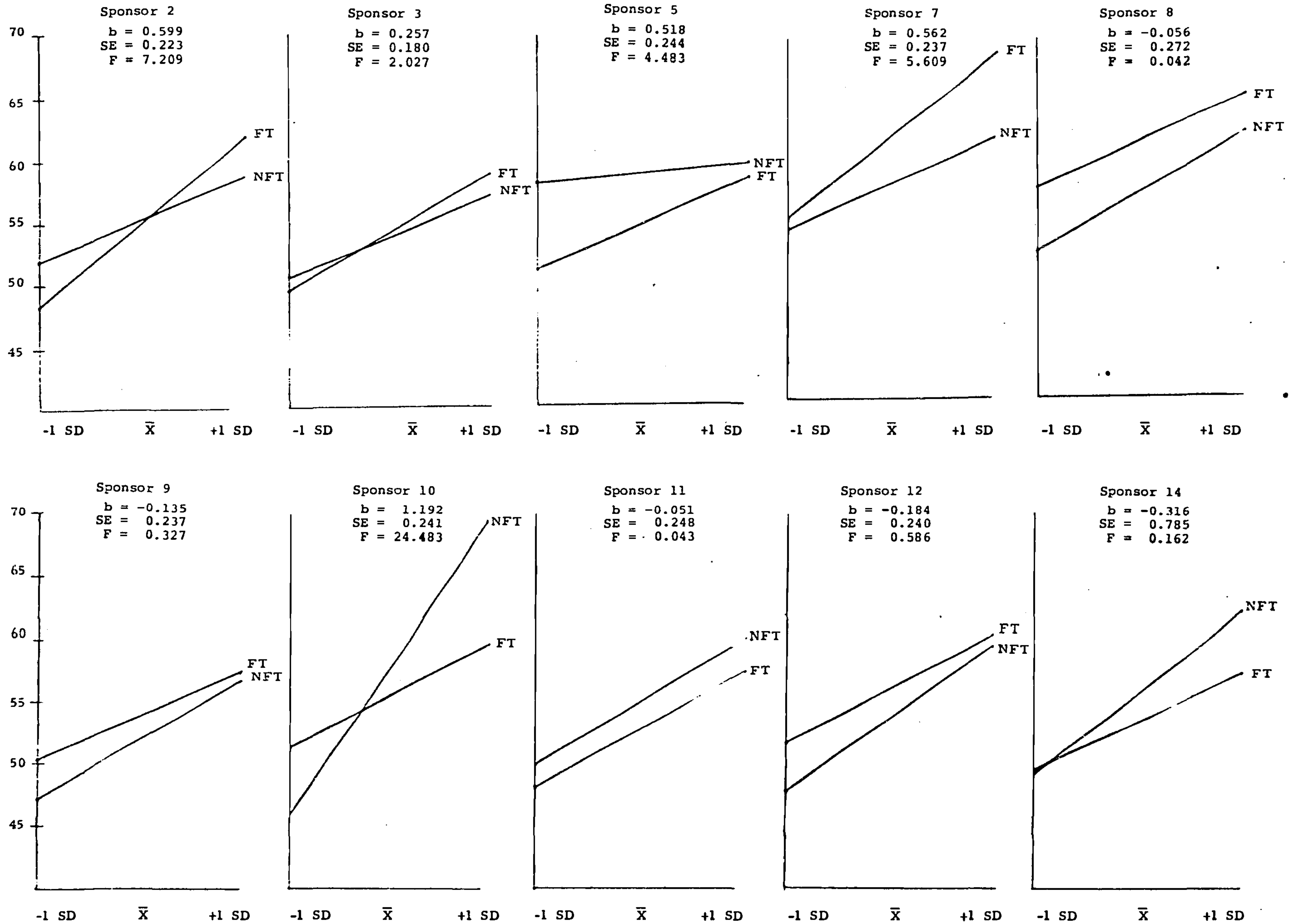


Figure VIII-2. Regressions of Spring WRAT on entry level for each Sponsor, by Follow Through and Non-Follow Through.

(d.f. = 1,353; $p < .05$)

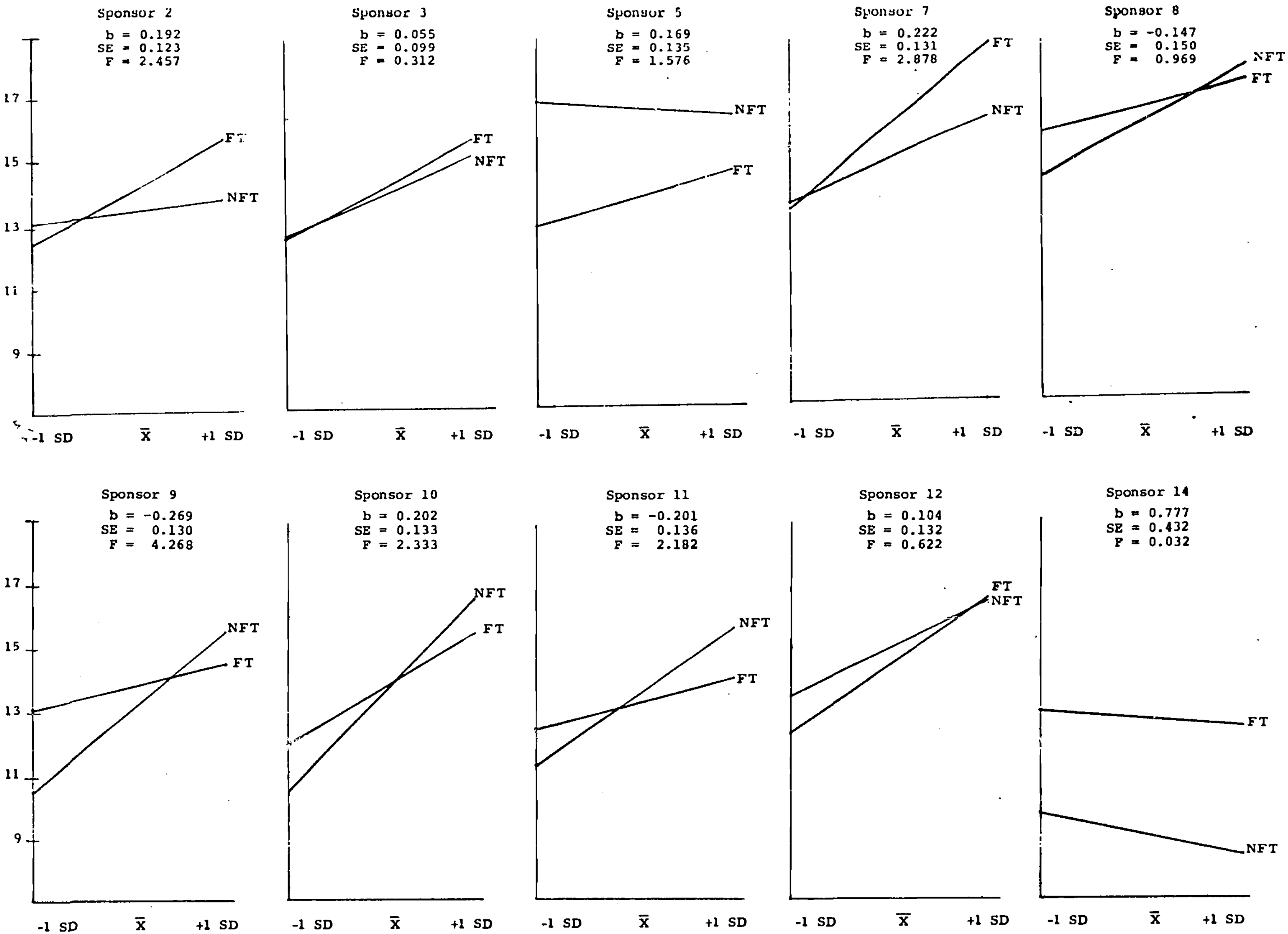


Figure VIII-3. Regressions of MAT Reading on entry level for each Sponsor, by Follow Through and Non-Follow Through.

(d.f. = 1,353; p < .05)

while the NFT coefficient is somewhat disparate. The latter coefficient is also based on fewer observations and is less representative.

The size of the difference between the FT and NFT regressions is sufficiently small to allow the use of initial achievement level as a covariable. However, the differences are substantial; the steeper slope in FT suggests that the higher entry level classes may benefit more from their experience than classes of lower initial ability. Tables VIII-4 and VIII-5 present the covariable-adjusted relationships for the WRAT and MAT reading subtest and indicate that the regression lines for FT and NFT are statistically parallel. Thus, there is no significant differential effect across entry level for these groups.

2.3.3 Sponsor 5: Bank Street College

Sponsor 5's FT and NFT classes have somewhat different distributions across initial achievement levels. The NFT group has a very small number of classes outside the range mean Fall WRAT $\pm$ 1 S.D. The regression for this group is thus more sensitive to minor variations in these few means. The groups have similar standard deviations but the NFT group has a 2.36 point advantage overall. The regression coefficient of the Spring WRAT on the Fall WRAT for the FT group is .59 and for the NFT group, .25. This difference is statistically significant ($F = 3.91$; $df = 1, 364$; $P < .05$). The coefficients indicate that in the FT group there is an increase of .59 points on class mean Spring WRAT for each point increase on the Fall WRAT and a .25 increase for the NFT classes.

The relatively flat regression in the NFT group is the result of the relatively large gain shown in the single NFT class with a lower-than-average initial achievement level and the relatively small gain shown in the classes with higher-than-average initial achievement levels. The classes with average initial achievement level ($\overline{X} \pm$ 1 S.D.) show an unadjusted average gain (UAG) of 22.22 points (Spring WRAT minus Fall WRAT), a value very similar to the overall UAG of 20.00 points. The single lower-than-average NFT class shows a UAG of 35.09 points while the two above-average classes showed a UAG of 16.33 points.

The use of initial achievement levels for the group results in some bias in the estimate of Sponsor 5's effects. The bias is due to the small number of observations and the extreme disparity of the low entry level NFT class. Tables VIII-4 and VIII-5 show the covariable-adjusted regressions in FT and NFT for the Spring WRAT and MAT reading subtest. The effect of the substantial gain in the low entry level NFT class and the relatively small gain in the high entry level NFT class can be seen in these figures. The regression line for NFT is flat, suggesting a uniform gain across entry level, while the regression line for the FT classes has a shallow, positive slope. These differences suggest, first of all, that the NFT group is atypical, having a substantially different regression from other NFT groups; and secondly, that the assessment of the Sponsor's effects are substantially biased both by the comparison with an atypical NFT group as well as by inappropriate covariable adjustment.

2.3.4 Sponsor 7: University of Oregon

Sponsor 7's FT and NFT classes have somewhat different distributions across initial achievement level. The NFT classes have a substantially larger standard deviation than the FT classes and the FT classes have a slight overall advantage, 1.39 points. The regression of the Spring WRAT on the Fall WRAT for the FT group is 1.24 while the regression in the NFT group is .65. The difference between these regressions is highly significant ($F = 7.16$; $df = 1, 364$; $P < .01$) and indicates a substantially steeper slope for the FT classes. Considering the distribution of classes in the FT and NFT groups, it is likely that these slopes are representative of true differences between the gains in the FT and NFT group. The regression of the covariable adjusted Spring WRAT and MAT reading subtest confirm this differential gain across initial achievement levels. The regressions of the Spring WRAT on the Fall WRAT and the MAT reading subtest on the Fall WRAT are shown in Figures VIII-2 and VIII-3. Both figures indicate substantially greater gains for higher initial achievement level in FT classes. That is, as initial achievement level increases, the advantage of the FT group increases. For both outcomes, the departure from parallel of the FT and NFT regression lines is statistically significant.

2.3.5 Sponsor 8: University of Kansas

Sponsor 8's FT and NFT classes have substantially different distributions across initial achievement levels, their standard deviations are very similar, and the FT group has a substantial overall advantage, 5.34 points. (See Tables VIII-6 and VIII-7.) The regression of Spring WRAT on Fall WRAT in the FT group is .72 and in the NFT group, .84. There is no statistical difference between these values. Thus, the Fall WRAT should be an appropriate adjuster for initial achievement differences, and there is no indication of the Sponsor's program having a differential effect across initial achievement level. Figures VIII-2 and VIII-3 indicate parallel regression lines in the FT/NFT groups for the covariable-adjusted WRAT and MAT reading outcomes, and confirm the absence of a differential effect.

2.3.6 Sponsor 9: High/Scope Educational Research Foundation

Sponsor 9's FT and NFT classes have very similar distributions across initial achievement levels and the standard deviations in the two groups are very similar, as are the overall initial achievement level means. (See Tables VIII-6 and VIII-7.) The regression of Spring WRAT on Fall WRAT reflects this comparability as well. The regression in the FT group is .73 and in the NFT group, .71. The Fall WRAT is suitable as a covariable for these groups since the regressions are almost identical. Also, there is no indication of the differential effectiveness of the program across initial achievement. (See Figures VIII-2 and VIII-3.)

2.3.7 Sponsor 10: University of Florida

Sponsor 10's FT and NFT classes have very different distributions across initial achievement level. The FT group has a 2.3 point overall advantage and the groups have similar standard deviations. As for distributional differences, the FT group has classes with higher-than-average initial achievement level where the NFT group has none. The NFT groups also has a fairly restricted range with a very small number of classes at low initial achievement levels. The regression of the Spring WRAT on the Fall WRAT in FT groups is .69 and in the NFT group, 2.06. This difference is highly significant ($F = 6.63$; $df = 1, 353$; $P < .01$). However, the steep regression

in the NFT group reflects the very small gain in low scoring NFT classes and is not necessarily representative of a difference in gains attributable to initial achievement level.

The three NFT classes with initial achievement means below average showed an unadjusted average gain (UAG) of 10.15 points (Spring WRAT minus Fall WRAT), while the classes with average initial achievement values showed a UAG of 24.11 points and the entire sample showed a UAG of 20.0 points. If this small gain is representative of Sponsor 10's low initial achievement level NFT classes, then his program is generally quite effective since his lower-than-average FT classes showed a UAG of 22.41 points. Tables VIII-4 and VIII-5 show the effect of the low gain NFT classes relative to the gain in FT. The regression in NFT is, however, somewhat deceiving in that there are no NFT classes in the range of initial achievement above the mean where NFT scores would exceed FT, according to the displayed regression.

Since the majority of Sponsor 10's NFT classes have initial achievement values within the category ± 1 S.D. about the mean on initial achievement, the use of initial achievement as a covariable will not bias the estimates of his effects substantially. However, there will be some bias reflecting the low gain, low achieving NFT classes. As a consequence, Sponsor 10's effects are likely to be slightly overestimated at the class level.

2.3.8 Sponsor 11: Educational Development Center

Sponsor 11's FT and NFT classes have similar distributions across initial achievement level but the FT group has an overall disadvantage of 1.98 points, and a substantially smaller standard deviation reflects a restricted range. The regression of the Spring WRAT on the Fall WRAT in the FT and NFT group are similar, .79 and .94, respectively. The smaller regression coefficient in the FT is likely due to the restriction of range on initial achievement level in this group. The initial achievement level covariable is appropriate for these groups and Sponsor 11's program shows no differential effect upon initial achievement level. (See Figures VIII-2 and VIII-3.)

2.3.9 Sponsor 12: University of Pittsburgh

Sponsor 12's FT and NFT classes have similar distribution across initial achievement level, although the FT group has a greater representation at higher initial achievement levels. The FT group also has a 1.4 overall advantage and a slightly higher standard deviation than the NFT group. The regression coefficient of the Spring WRAT on the Fall WRAT for the FT group is .79 and for the NFT group .94. These coefficients are comparable and permit the use of the Fall WRAT as a covariable. Again there is no differential effect across initial achievement level. (See Figures VIII-2 and VIII-3.)

2.3.10 Sponsor 14: Southwest Educational Development Laboratory

Sponsor 14's FT and NFT classes have similar distribution and the FT group has an overall advantage of 1.8 points. The standard deviation of the FT group is, however, nearly two times as large as the NFT group. The regression coefficient for the Spring WRAT on the Fall WRAT for the FT group is .72 and for the NFT group, 1.2. In spite of the large size of the differences between these groups, the difference is not statistically significant, reflecting the instability of the estimates. The NFT group has both too small a number of classes and too restricted a range for an accurate estimate of regression. The use of initial achievement level as a covariable will result in some bias. Tables VIII-4 and VIII-5 indicate no difference between FT and NFT regressions, and attest to the instability of regression.

2.4.0 DISCUSSION

The results indicate heterogeneity of regression only for the Spring WRAT and only for some Sponsors. Furthermore, for two of the four Sponsors for whom heterogeneity is found, the heterogeneity can be accounted for in terms of the distribution of the sample of classes across initial achievement level in the Sponsor NFT group. For Sponsors 5 and 10 the NFT groups show regressions of Spring WRAT on Fall WRAT that are substantially different from their FT groups as well as from all other groups in this study. In both cases, the importance of the disparate regression is brought into question by the fact that these groups contain a very restricted representation across initial achievement level. For these Sponsors the NFT regression is

highly unstable. The instability of the regression suggests that the heterogeneity is spurious. For the other two Sponsors, Sponsor 2 and 7, the heterogeneity is obtained because the FT group has a substantially steeper regression than the NFT group, which displays a regression that is very similar to other NFT groups in the sample. For both Sponsors the steeper slope in FT indicates a greater gain as initial achievement level increases. For Sponsor 2, the FT classes with higher-than-average initial achievement level gain more than NFT classes with a comparable initial achievement level. Classes of average or below average achievement level do not fare as well. For classes of average initial achievement level FT and NFT produce equal outcome scores and for classes of lower-than-average initial achievement level, the NFT classes exceed the FT classes. For Sponsor 7, a substantially different picture is obtained. Sponsor 7's FT classes exceed the NFT classes across all levels of initial achievement score. However, the advantage of FT increases with an increase in initial achievement level.

Possibly the most important aspect of the results is the general inadequacy of the sample size for assessing the relationship between initial achievement level and outcomes. For most Sponsors, the number of classes in FT and NFT is minimally appropriate for the estimation of regression effects and much of the heterogeneity can be accounted for in terms of the inappropriateness of the distributions of classes across initial achievement level.

3.0 CHILD LEVEL PRESCHOOL STUDY

3.1.0 INTRODUCTION

The relationship of a child's preschool experiences to his performance in Project Follow Through is of interest both to educational policy planners and to researchers because of its implications for the child's long-term educational development. The concept of compensatory early education is predicated on the assumption that a child's ultimate success in school can be enhanced by providing a foundation of academic skills, learning abilities and positive school attitudes at the outset of the child's school career. This philosophy, of course, gave rise to Project Head Start; however, the continuing value to the elementary school child of the educational benefits derived from Head Start has yet to be conclusively demonstrated.

In a recent review of a number of longitudinal Head Start studies, Beller (1973) noted that children from lower socio-economic levels with preschool experience tend to achieve higher levels of academic, cognitive and/or affective functioning than their non-preschool peers. However, comparable non-preschool children often catch up to the school performance of the preschool graduates by the end of the second or third grade. This pattern was found for IQ scores (Weikart, 1970), achievement test scores (Gray and Klaus, 1970), and self-concept measures (Gray and Klaus, 1970). One of the goals of Project Follow Through is to interrupt this pattern, so that whatever advantages are acquired in Head Start will be maintained in elementary school.

Past research suggests a number of factors which influence the elementary school performance of preschool graduates. First, consider the elementary school environment. Hyman and Kliman (1967) noted that Head Start children who attended elementary schools located in middle income neighborhoods maintained their academic advantage while Head Start graduates who entered schools servicing lower income neighborhoods lost their initial advantage over their non-Head Start comparison group. In another study of Head Start graduates in a large city elementary school system, Wolff and Stein (1966) reported that some kindergarten teachers "extinguished" the questioning and exploratory behavior of Head Start children in a manner which may have inhibited

academic growth. Clearly, the impact of preschool experience can be significantly influenced by subsequent primary school experiences.

A second broad factor affecting the elementary school child's performance is the nature of the preschool program which the child experienced. In Beller's (1973) extensive review of preschool programs, he discussed several studies designed to compare the gains of children who had experienced different types of preschool and Head Start curricula. In the Karnes (1969) study, a behaviorally structured program, a direct ameliorative training, a Montessori, and a traditional preschool program were compared. By the end of preschool the behavioral and direct training programs had produced greater gains in specific academic areas than the other two programs. By the end of first grade, however, there were no significant differences among the four groups on Stanford-Binet IQ scores, although the behavioral and direct training group had higher achievement test scores. These higher scores were attributed in part to the greater supplementary training received by those two groups in kindergarten. A second study by Weikart (1970) compared behavioral, cognitively oriented, and traditionally oriented preschool programs. While all three groups demonstrated equally high gains in IQ after a two-year preschool experience, by the end of second grade there was a distinct trend toward lower achievement test scores in the behavioral preschool group. These and other comparative preschool studies suggest that different types of preschool programs have different kinds of effects on the performance of children, both because of the theoretical model underlying the curriculum and because of factors influencing teacher training and supervision and other implementation issues.

Still a third factor influencing the relationship of a child's preschool to his elementary school experience are the motivational and affective orientations and the developmental history of the child. Beller (1972) reported an apparent "timing" effect: children who entered preschool two years before first grade demonstrated significant IQ gains which were maintained through fourth grade. Children who entered kindergarten with about the same entering scores as the first group gained slightly less than the first group but maintained these gains through fourth grade. Children in a third group entered first

grade with no preschool experience, and demonstrated no IQ gains. Thus, the earlier the children started school the greater and the more persistent were their IQ gains. In this same study, Beller noted that for children high in autonomous achievement striving, time of school entry made no difference in their later IQ levels, whereas the low autonomous achievement strivers demonstrated a definite drop in IQ levels by the end of first grade.

This discussion has briefly summarized three general aspects of a child's experience which influence the relationship between preschool and elementary school performance: the type of preschool experience; the context and content of the elementary school experience; and the affective, motivational and developmental characteristics of the child. The present study focuses on one of these dimensions: the nature of the preschool experience. We seek to examine differences among Sponsors in the manner in which children with different types of preschool experience respond to the kindergarten year in Project Follow Through.

Three categories of preschool experience were used in this study: Head Start attendance; other preschool attendance; and no preschool attendance. Children in the first group were enrolled in federally-funded Head Start programs; however, there is no information on the educational content of any of these programs. Children in the second group attended other non-Head Start preschool programs. Descriptions of these programs and their similarities to the various Head Start programs are also lacking in current data. Children in the third group had remained home prior to kindergarten entry.

It is important to note that while the ten Sponsors involved in this study represent a wide range of theoretical approaches to education, we do not yet have data which describe in detail the educational contents and processes of these models, nor do we have information concerning the relationships of the various programs to the school, community and broader social environments in which the models are implemented. In the absence of detailed descriptions of the types of preschool programs and the Sponsors' models implementation programs, this study cannot provide explanatory relationships in terms of the psychodynamics and social contexts affecting school performance. More detailed explanatory

analyses incorporating some of these factors will be conducted when more complete data become available from the Planned Variation Head Start Study in the coming year. This study does, however, allow an initial assessment of differential Sponsor effects between Head Start *per se* as a policy-defined group and groups with other types of preschool experience.

3.2.0 METHOD

3.2.1 Subjects

The subjects for this study were drawn from Cohort III kindergarten. Only those subjects with a complete pretest and posttest battery from the Fall 1971 and Spring 1972 testings as well as a parent interview and teacher questionnaire were included. Of the approximately 21,000 kindergarten children tested in Fall 1971, and 11,000 tested in Spring 1972, approximately 5,000 met these selection criteria. Since ethnic types other than Blacks or Whites were too sparsely distributed across the Sponsors to allow adequate analyses, only Black and White children were selected. In addition, children who had received both Head Start and some other preschool experience were eliminated from the analytic sample. Ten Sponsors with sufficient subjects for adequate analysis remained after these selection considerations. The resulting distribution of 3,580 subjects across ten Sponsors is indicated in Table VIII-8. Included for descriptive purposes are the respective pretest means on the Wide Range Achievement Test (WRAT) scores from Fall 1971, the adjusted income index and the proportion of mothers with at least a high school diploma.

3.2.2 Measures

The outcome measures analyses in the child studies include academic achievement tests and measures of motivational orientation taken from the Spring 1972 kindergarten test battery, and a measure of absence indicated by the number of days the child missed school. These measures are described in detail in Appendix A. Briefly, they are:

- Wide Range Achievement Test (WRAT)
- Peabody Pictures Vocabulary Test (PPVT)
- Metropolitan Achievement Tests

    Reading (MAT-Reading)

N = 3580			SPONSORS 2	3	5	7	8	9	10	11	12	14	GROUP MEANS
FT	Head Start	N = 1166	177	131	95	111	180	93	88	106	138	47	
		WRAT	37.23	30.95	34.73	37.46	36.03	34.58	31.30	33.60	34.58	31.43	36.62
		Adjusted Income	9.10	8.65	10.15	7.48	8.72	8.59	8.31	10.52	9.26	8.60	8.95
		Mother's Education	0.49	0.50	0.48	0.41	0.43	0.54	0.38	0.43	0.64	0.47	0.48
	Preschool	N = 336	36	52	29	10	29	21	53	65	37	4	
		WRAT	42.08	44.23	42.41	36.30	41.24	37.14	38.98	37.99	47.65	33.75	41.12
		Adjusted Income	14.78	16.29	13.62	8.90	11.07	11.71	13.43	13.88	17.92	10.50	14.13
		Mother's Education	0.75	0.87	0.76	0.90	0.48	0.86	0.68	0.63	0.89	0.50	0.74
	No Preschool	N = 585	47	101	79	13	72	39	53	65	71	21	
		WRAT	35.68	38.02	31.57	36.62	34.86	30.49	31.88	32.95	42.07	32.48	34.96
		Adjusted Income	13.04	15.32	13.39	6.54	10.85	11.41	13.51	12.94	14.06	13.19	13.13
		Mother's Education	0.70	0.68	0.49	0.62	0.44	0.56	0.47	0.57	0.78	0.48	0.58
NFT	Head Start	N = 329	83	50	9	33	53	30	10	13	34	14	
		WRAT	35.33	33.24	27.89	32.70	29.64	32.50	28.20	34.46	39.24	31.07	33.34
		Adjusted Income	8.40	10.24	10.67	10.30	7.94	7.63	7.40	10.31	12.03	8.14	9.20
		Mother's Education	0.46	0.50	0.56	0.64	0.36	0.33	0.40	0.46	0.62	0.21	0.46
	Preschool	N = 339	57	88	28	27	13	32	23	39	27	5	
		WRAT	39.53	45.74	40.82	44.15	35.54	40.91	35.91	39.23	41.11	37.40	41.41
		Adjusted Income	14.74	19.40	15.75	15.93	14.31	18.97	14.74	15.05	16.33	16.40	16.70
		Mother's Education	0.58	0.90	0.82	0.59	0.69	0.88	0.74	0.62	0.82	1.00	0.76
	No Preschool	N = 825	143	118	94	59	41	79	68	88	93	42	
		WRAT	36.92	39.74	37.80	34.27	27.98	29.60	35.65	37.28	36.42	28.62	35.54
		Adjusted Income	13.82	15.61	15.47	13.34	11.81	13.14	15.63	14.19	15.50	12.33	14.37
		Mother's Education	0.64	0.72	0.62	0.51	0.37	0.56	0.54	0.66	0.84	0.43	0.62

TABLE VIII - 8

Distribution of subjects for Head Start by Sponsor by FT/NFT analysis with entering WRAT, Poverty Level, and Mother's Education group means

Numbers (MAT-Arithmetic)
Listening for Sounds (MAT-Listening)

- Gumpgookies
- Locus of Control

  Locus (positive)
  Locus (negative)

- Absence

The covariates are discussed in detail in Chapter IV. They include characteristics describing the child, parent and home, classroom, and community as outlined briefly below:

- Fall WRAT. This test served as the pretest covariate for all measures except the PPVT, whice used the Fall PPVT as the pretest covariate.
- Adjusted income. A measure reflecting family income adjusted for family size and urban/rural residence.
- Mother's education. Categorized as high school diploma or greater versus less than high school diploma.
- Years at current address.
- School receptivity. As perceived by the parents.
- Parent-school involvement. As reported by the parents.
- Teacher's years of education.
- Teacher's years of teaching experience.
- Percentage of White pupils in the classroom.
- City size. Coded on a scale of 1 through 4 for populations ranging from under 10,000 to over 200,000.

3.2.3 Analytic Plan

The data from this study were analyzed with the multiple regression equivalent of a three-factor, fully crossed analysis of covariance technique. As outlined in Table VIII-8, there were ten levels of the Sponsor factor, two levels of the FT/NFT factor, and three levels of preschool experience: Head Start; other preschool; and no preschool. All three factors were effects coded for the multiple regression analysis in the manner described by Cohen (1968). Correlations with the WRAT pretest covariate were adjusted in accordance with the reliability of that test, using the adjustment procedure suggested by Porter (1973).

The F-ratios for the three-way interaction terms on each outcome variable discussed in the Results section are computed as follows:

$$F = \frac{R^2_{y.cov,\ ABCDEFG} - R^2_{y.cov,ABCDEF}}{1 - R^2_{y.cov,\ ABCDEFG}} \times \frac{N - 69 - 1}{18}$$

The components of variance in the F-ratio are defined as follows:

Analytic Model Components		df
cov - covariates listed above		10
A preschool experience		2
B Sponsor = 2,3,5,7,8,9,10,11,12,14		9
C FT/NFT		1
D Preschool by Sponsor		18
E Preschool by FT/NFT		2
F Sponsor by FT/NFT		9
G Preschool by Sponsor by FT/NFT		18
y Outcome variable	TOTAL =	69

Details of this analytic technique are given in the Methodology Appendix of this report.

3.3.0 RESULTS

The data in Table VIII-8, although not treated analytically here, indicate that the Head Start (HS) and no preschool (NPS) groups had equivalent entering achievement scores, while the other preschool groups (PS) had distinctly higher entering scores. The lower levels of income and mothers' education in the HS group, however, would lead us to expect lower entering achievement scores in that group. Thus, it is likely that the Head Start children had higher entry scores than they would have had without their preschool attendance. Similarly, although the income levels of the PS and NPS groups are approximately equal, the entering achievement level of the PS group is higher than that of the NPS group. This pattern suggests that the preschool and Head Start experiences may have provided some academic benefits not available to the NPS group. A more detailed analytic statement of this potential effect, however, must await the analysis of data (available this coming year) on the status of the groups prior to their preschool experience.

The three-way interaction of preschool group by Sponsor by FT/NFT showed significant effects on the WRAT, MAT-Arithmetic and MAT-Reading

tests, but no significant effects on the motivational or absence measures. The significant interactions suggest that in the achievement areas FT/NFT differences vary across Sponsors and among the three different categories of preschool experience. Therefore, child-level analyses which combine the three preschool groups (such as the Sponsor by FT/NFT interactions) or combine the Sponsors (such as the preschool by FT/NFT interactions) are inappropriate and misleading.

In order to compare the FT/NFT differences among the three preschool groups for each separate Sponsor, the Spring scores of each group, statistically adjusted for all covariates, were plotted for every Sponsor in Figure VIII-4 for WRAT scores, in Figure VIII-5 for MAT-Arithmetic, and in Figure VIII-6 for MAT-Reading. The plots are arranged so that lines sloping upward to the right indicate higher scores for FT. Since there is inevitably some variation around the group means as plotted, the Sponsors with contrasts an F ratio of 2.0 or greater are noted by an asterisk. The Sponsors without an asterisk may only reflect trends within the general limits of error variability.

The three-way interactions in Figures VIII-4 through VIII-6 show several different types of patterns. Some Sponsors appear to produce roughly equivalent effects in all three preschool groups, for example the University of Arizona (3) and the University of Oregon (7) on all three measures, the University of Kansas (8) on WRAT and MAT-Arithmetic, and High/Scope (9) on WRAT and MAT-Reading. Some Sponsors appear to produce greater FT effects in the Head Start group than the no preschool group: for example Educational Development Center (11) in MAT-Arithmetic and MAT-Reading, Far West Laboratory (2) in WRAT and MAT-Arithmetic, Southwest Laboratory (14) in MAT-Reading, and Bank Street College (5) in MAT-Arithmetic. Still a third pattern of FT effects--lower for the Head Start group than the no preschool group--is indicated for the University of Pittsburg (12) on MAT-Arithmetic and Bank Street College (5) on MAT-Reading and WRAT. In general, these interactions do not provide patterns which characterize different theoretical models of education consistently across various Sponsors and outcomes.

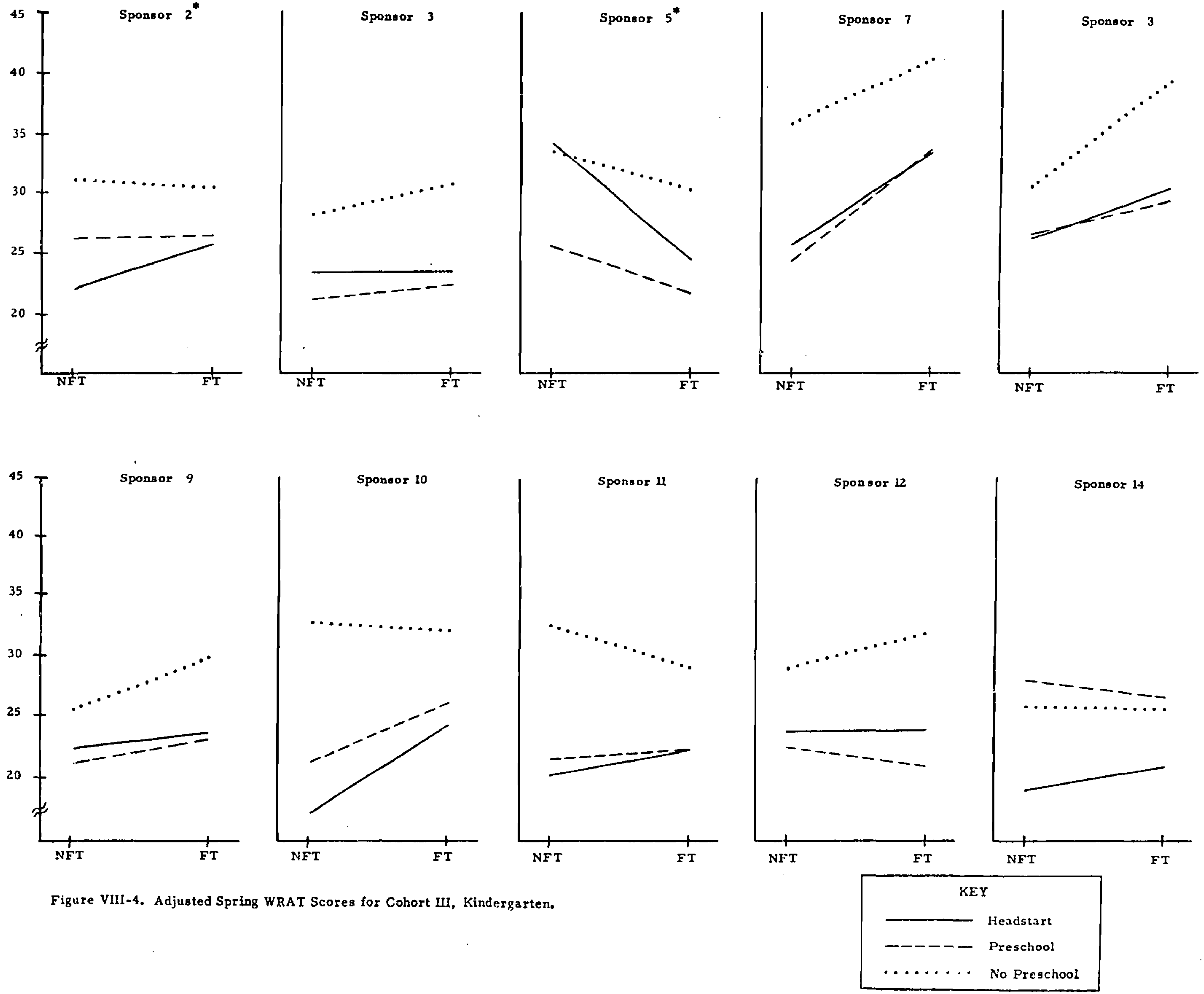


Figure VIII-4. Adjusted Spring WRAT Scores for Cohort III, Kindergarten.

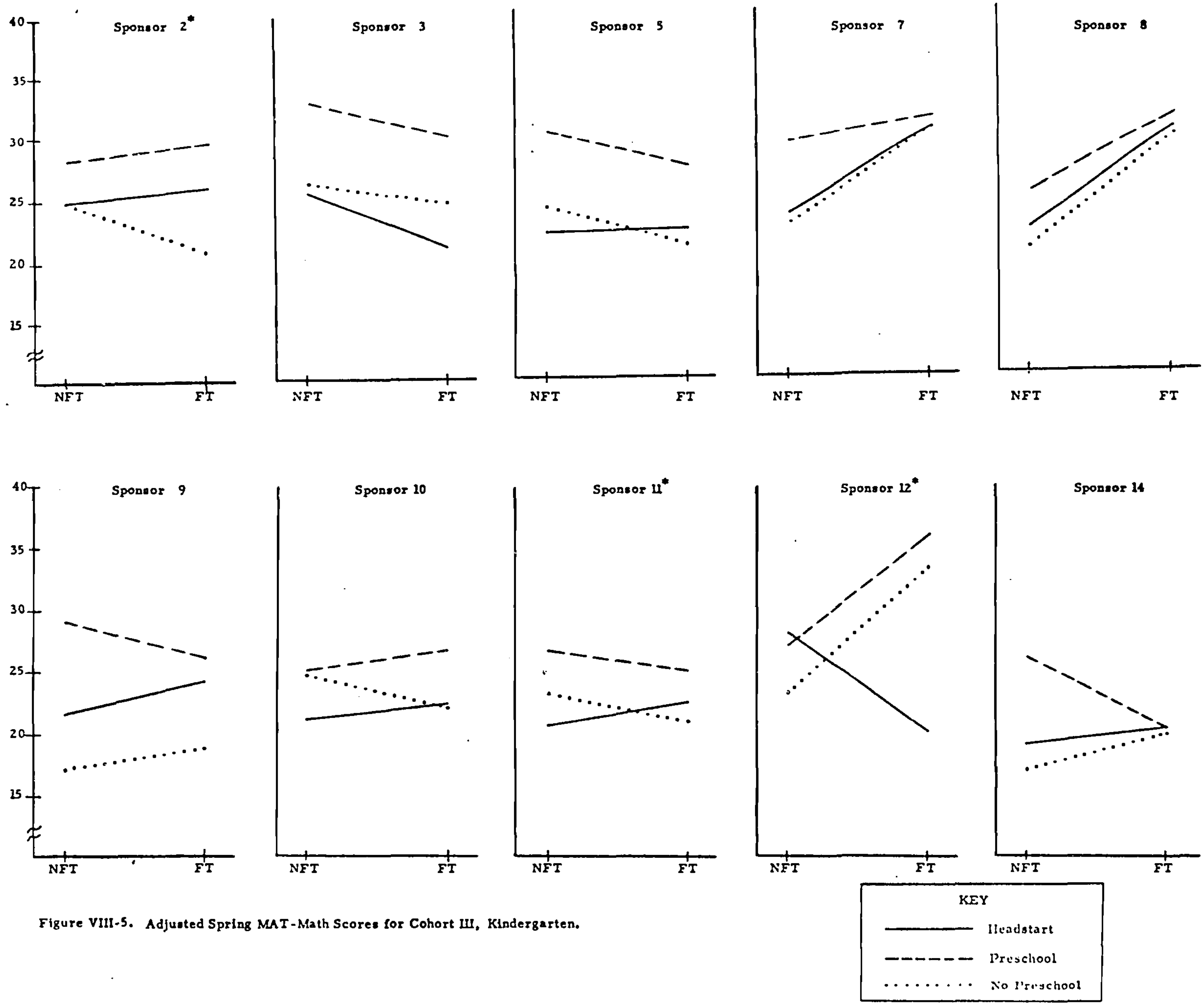


Figure VIII-5. Adjusted Spring MAT-Math Scores for Cohort III, Kindergarten.

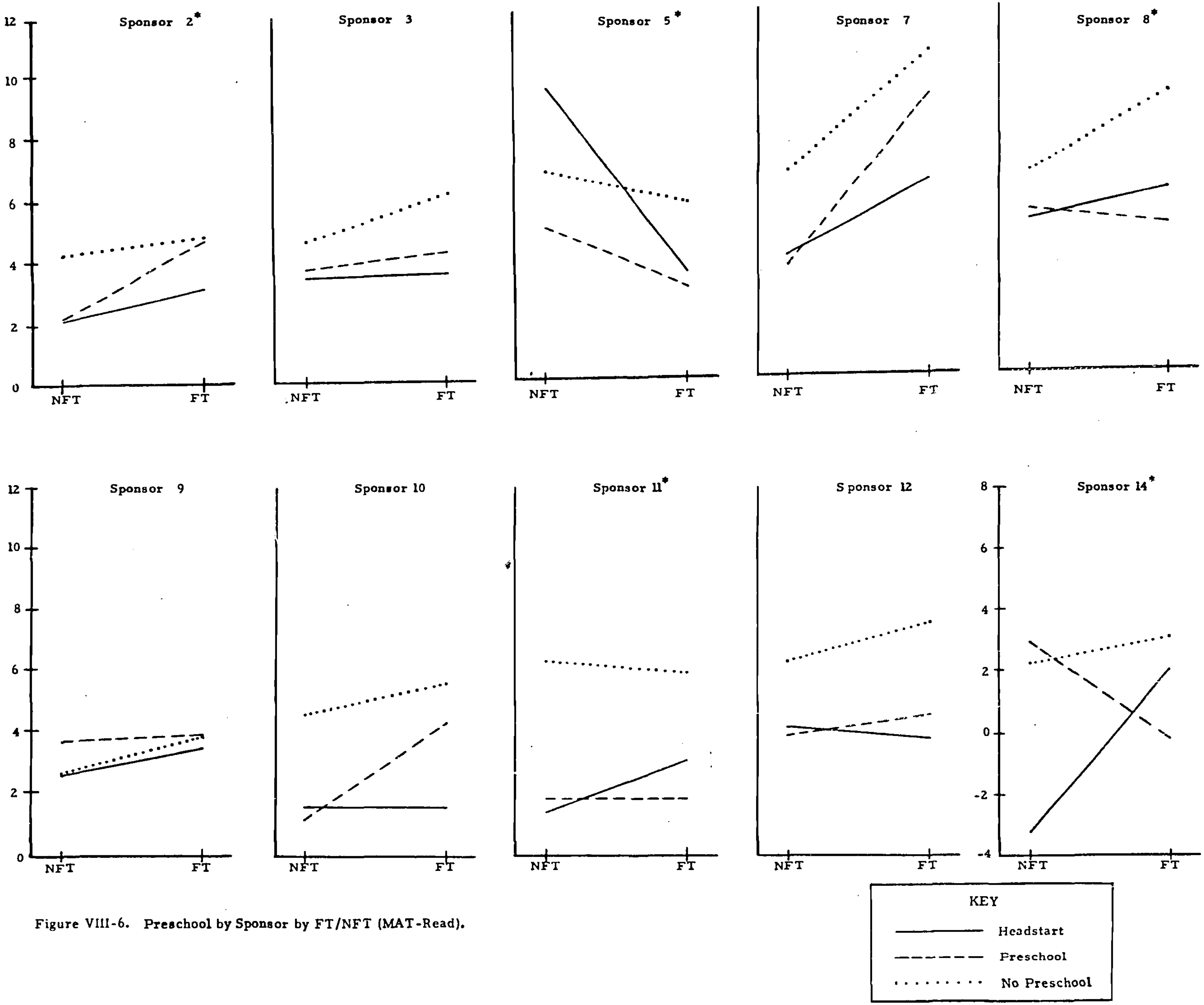


Figure VIII-6. **Preschool by Sponsor by FT/NFT (MAT-Read).**

3.4.0 DISCUSSION

A brief review of the research in early education suggests two patterns: first, that children attending a preschool program may enter kindergarten or first grade with an advantage in academic areas over comparable children who did not attend preschool; and second, that this initial difference between such groups of children tends to fade as the children progress toward third grade. The types of effects of preschool and their persistence into elementary school seem to be related to the nature of the preschool experience, the context and content of the elementary school and the developmental characteristics of the child. Descriptive data from this study, consistent with the first of these patterns, suggests some entering achievement advantages attributable to preschool attendance. The purpose of this study is related directly to the second of these patterns: To determine the extent to which Follow Through Sponsors are able to build on the Head Start experiences of children in a manner which avoids the loss of these entering achievement advantages in elementary school.

Relative to the second point noted above, Follow Through Sponsors would have altered the trends of past research if they had produced equivalent Follow Through effects in children from all types of preschool experience, or if they had demonstrated greater Follow Through effects in the Head Start group than the no-preschool group. The results of this study indicate that in academic achievement areas a number of Sponsors did in fact produce these patterns, while a few Sponsors demonstrated effects indicating that Head Start children gained less in Follow Through than in non-Follow Through schools. In the motivational areas and in absence there were no significant effects related to Sponsors or preschool groups. Although no patterns emerge which would support generalizations across theoretical models, it appears that a number of different Sponsors are having some success in Project Follow Through's attempt to build on previous Head Start experiences in achievement areas.

In the motivational areas a parsimonious explanation for the lack of significant effects is that the lower reliability of these measures allows for less meaningful variance and relatively more error variance

in the analytic model. In fact, the total variance accounted for by the hypotheses and covariates in the motivational areas was much lower than in the achievement areas (Table A VIII-1, Appendix A).

What remains unspecified in the current study is the nature of the Head Start experiences and their congruence with the Follow Through experience for the children in the various Sponsors. We know that the types of Head Start schools are as varied as the types of Follow Through models, and that an educational experience of one type may either facilitate or inhibit the adjustment to a classroom of a quite different type for certain kinds of children. Because the affective orientation of the child and the instructional dynamics of the classroom have not been specified, much information is lost that otherwise might help to explain the obtained differences among the three preschool groups in the various Sponsors. The present study does demonstrate that when Head Start is considered on the whole in a policy-defined sense, a number of Sponsors are significantly building on the gains of Head Start children. A more detailed assessment of the relationship of different types of Head Start to different types of Follow Through models must await the analysis of the Planned Variation Head Start data currently under study.

4.0 CHILD ETHNICITY STUDY

4.1.0 INTRODUCTION

Among the several child characteristics that may be explored as predictors of differential Sponsor FT effects is the ethnicity of the child. Child ethnicity is of analytic interest because it is a general proxy variable representing a range of social, economic and psychodynamic forces acting differently upon Black, White and other ethnic types of children.

Stein (1971), Cohen (1968) and others have described how social and economic pressures may operate in some environments to support the existing structure and to resist change of an educational system which perennially produces, in a large number of children, achievement scores falling far below grade level norms. The fact that this segment of the school population consists mostly of non-White ethnic types does not explain the processes underlying such endemic school failure, but it does suggest that the results of schooling simply are not equal for different ethnic groups in our society. The relationships of ethnic differences in school performance to the social environment is outlined in part by Coleman (1966) who indicated that Blacks in integrated classes tend to perform better than Blacks in segregated classes, while the school performance of Whites in integrated classes was not significantly different from that of Whites in segregated classes. One might infer broadly from these results that ethnic differences in school performance may be responsive to the social context of the classroom.

Other findings in the Coleman Report suggest that the socio-economic background of the student's family is a stronger determinant of school performance than are differences in the character of the schools. In the sample available for analysis in this report, the Black parents as a group quite consistently have lower levels of income than do the White parents within each Sponsor (see Table VIII-9). The same pattern is true, with few exceptions, for the levels of mothers' education for the Blacks and Whites within each Sponsor (Table VIII-9). These consistent ethnic differences in socio-economic level may reflect, in part, the operations of pervasive social and financial pressures which favor the White majority. Inasmuch as these forces affect the home

N = 3830			SPONSORS										GROUP MEANS
			2	3	5	7	8	9	10	11	12	14	
FT	Black	N = 1219	169	139	64	120	256	129	101	139	68	34	
		WRAT	35.47	34.38	40.63	37.11	35.48	34.35	34.37	34.39	31.50	31.00	35.10
		Adjusted Income	9.89	9.44	10.58	7.55	8.31	8.87	9.29	10.99	8.02	8.00	9.12
		Mother's Education	0.59	0.55	0.64	0.51	0.46	0.60	0.49	0.45	0.69	0.38	0.53
	White	N = 1070	133	178	148	33	50	44	142	111	190	41	
		WRAT	39.85	37.19	32.17	40.39	40.64	33.23	32.78	34.63	41.11	32.39	36.58
		Adjusted Income	11.12	14.17	12.28	7.33	14.62	12.67	12.77	13.47	13.00	12.29	12.71
		Mother's Education	0.59	0.67	0.47	0.21	0.48	0.64	0.46	0.62	0.72	0.59	0.58
NFT	Black	N = 507	114	68	10	84	86	53	25	33	16	18	
		WRAT	33.26	32.69	36.90	33.69	29.11	33.38	29.88	31.94	28.81	25.83	31.98
		Adjusted Income	8.02	10.66	10.90	12.21	9.57	11.23	8.48	11.76	13.75	6.83	10.13
		Mother's Education	0.46	0.50	0.70	0.55	0.41	0.49	0.56	0.30	0.69	0.17	0.47
	White	N = 1034	189	197	121	35	24	91	76	114	144	43	
		WRAT	38.94	42.70	37.80	41.80	31.96	32.13	36.65	39.03	38.83	31.61	38.38
		Adjusted Income	14.58	17.32	15.55	15.17	12.42	14.36	16.63	14.53	15.13	13.74	15.35
		Mother's Education	0.64	0.82	0.65	0.60	0.42	0.62	0.58	0.70	0.79	0.54	0.69

TABLE VIII - 9

Distribution of subjects for ethnic type by Sponsor by FT/NFT analysis with entering WRAT, Poverty Level, and Mother's Education group means

and school environment of the child, they also differentially affect the entering achievement levels and subsequent school performance and adjustment of children of different ethnic groups. Analysis of Sponsor effects which do not consider these factors may not reveal the full strength of Follow Through as an intervention program. Consequently, we have examined Sponsors' FT/NFT contrasts as they are mediated by the ethnicity of the children involved.

4.2.0 METHOD

4.2.1 Subjects

The subjects for this study were drawn from the Cohort III kindergarten group according to the same criteria discusses in the Head Start child study. Only subjects with complete Fall 1971 and Spring 1972 test batteries, as well as parent interview and teacher questionnaire data were included. Since ethnic types other than Blacks or Whites were too sparsely distributed across the Sponsors to permit adequate analysis, only Blacks and Whites were included in this study. The resulting distribution of 3,830 subjects across the Sponsor, FT/NFT and ethnicity groups is indicated in Table VIII-9, along with the group means on entering WRAT scores, adjusted income index and mean proportion of mothers with at least a high school diploma. The number of subjects in this study is slightly larger than in the Head Start study because children were included here who had both Head Start and other preschool attendance. Note that the uneven distribution of subjects within Sponsor groups (in particular, Sponsors 5 and 12) may adversely affect the representativeness of any one Sponsor's effects across a range of sites.

4.2.2 Measures

The same nine outocmes reported in the Head Start study were analysed here. The reader is referred to Appendix A for a complete discussion of these measures. They are:

- Wide Range Achievement Test (WRAT)
- Peabody Pictures Vocabulary Test (PPVT)
- Metropolitan Achievement Tests

  Reading (MAT-Reading)
  Numbers (MAT-Arithmetic)
  Listening for Sounds (MAT-Listening)

- Gumpgookies
- Locus of Control

  Locus (positive)
  Locus (negative)
- Absence

The covariates for this study are identical to those for the Head Start study with one exception. Preschool experience, coded here as Head Start or other preschool versus no preschool, was included here as a covariate. Refer to Chapter IV for a full discussion of the covariates. Briefly they are:

- Fall WRAT
- Fall PPVT (used only for the PPVT outcome)
- Preschool experience
- Adjusted income
- Mother's education
- Years at current address
- School receptivity
- Parent-school involvement
- Teacher's years of education
- Teacher's years of teaching experience
- Percentage of White pupils in the classroom
- City size

4.2.3 Analytic Plan

The data from this study were analyzed with the multiple regression equivalent of a three-factor, fully crossed analysis of covariance. There were ten levels of the Sponsor factor, two levels of the FT/NFT factor, and two levels of ethnicity as outlined in Table VIII-9. Details of this analysis are discussed in the Methodology Appendix of this report.

The F-ratios for the three-way interaction terms discussed in the Results section are computed as follows:

$$F = \frac{R^2_{Y.cov,\ ABCDEFG} - R^2_{Y.cov,ABCDEF}}{1 - R^2_{Y.cov,ABCDEFG}} \times \frac{N - 50 - 1}{9}$$

The components of variance in the F-ratio are defined as follows:

Analytic model components		df
cov - Covariates lsited above		11
A	Ethnicity	1
B	Sponsors	9
C	FT/NFT	1
D	Ethnicity by Sponsors	9
E	Ethnicity by FT/NFT	1
F	Sponsors by FT/NFT	9
G	Ethnicity by Sponsors by FT/NFT	9
		50

Y Outcome variable

4.3.0 RESULTS

The purpose of this study was to assess the differential effects of Sponsors working with Black or White children. These effects may be assessed by examining the three-way interaction terms of ethnicity by Sponsor by FT/NFT. The F ratios for these interactions attained statistical significance on four outcomes (as indicated in Table A VIII-2 of Appendix A): WRAT, MAT-Arithmetic, MAT-Listening, and PPVT. No significant three-way interactions were produced on the motivational measures or on absence. Greater errors of measurement in the affective tests produced so much error variance in the analytic model variables that little true variance remained for the effects of interest. The interactions in the achievement areas suggest that Black children and White children do not respond in the same way to all Sponsors' Follow Through programs. In some Sponsors Blacks gain more than Whites relative to their respective NFT groups. In other Sponsors the converse is true; in still others both groups gain equally with respect to their NFT peers. Furthermore, these significant three-way interactions indicate that for these achievement outcomes it is inappropriate to combine different groups of children across either Sponsors or ethnicity to look at two-way interactions.

In order to study these different Sponsor effects more clearly, the FT/NFT contrasts within each Sponsor for Blacks and Whites separately were plotted for the four achievement outcomes in Figures VIII-7 through VIII-10. The Spring outcomes, adjusted for all covariates, are displayed such that an upward slope reflects a positive contrast of FT to NFT. Considering first the WRAT outcomes displayed in Figure VIII-7, it is

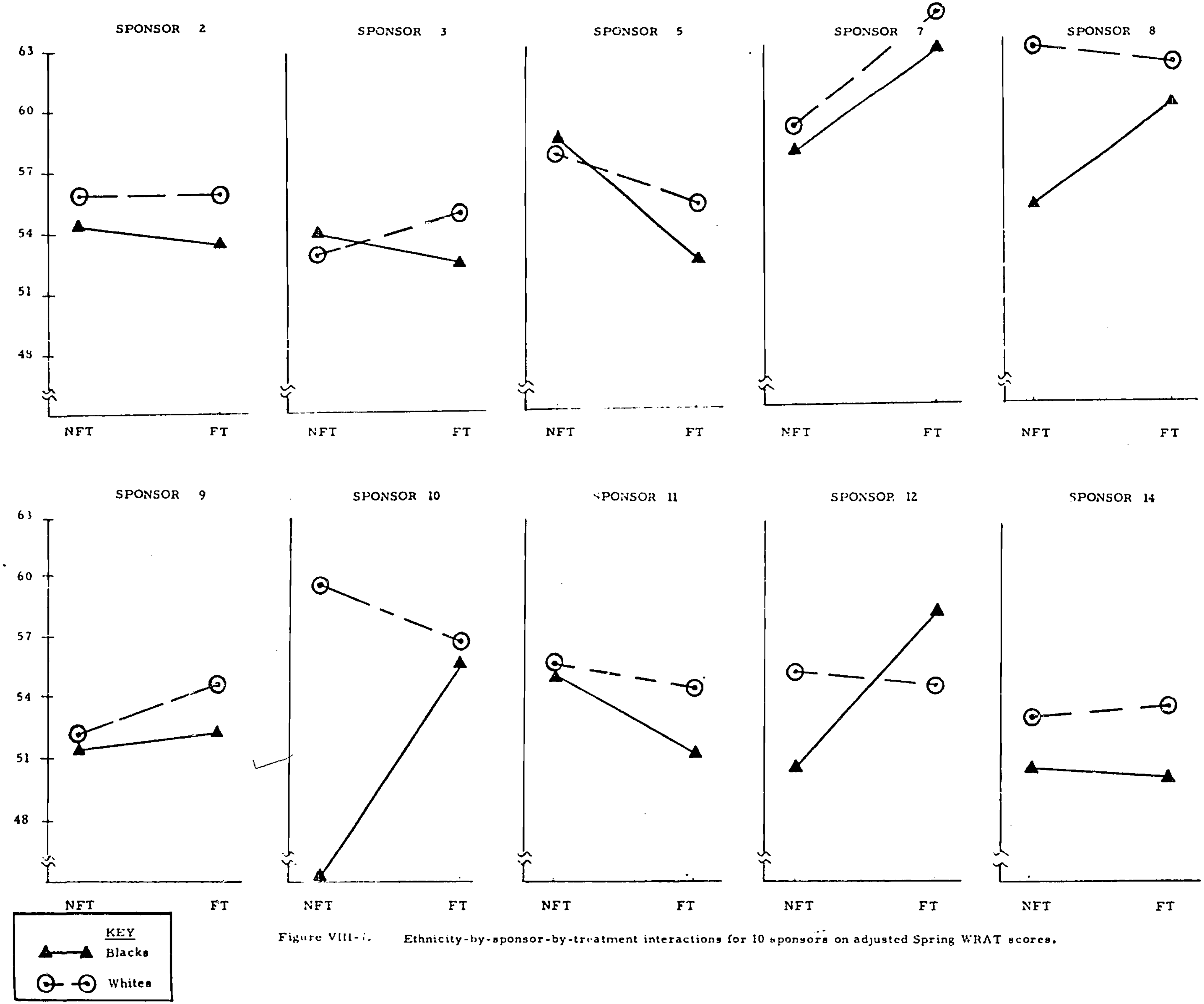


Figure VIII-7. Ethnicity-by-sponsor-by-treatment interactions for 10 sponsors on adjusted Spring WRAT scores.

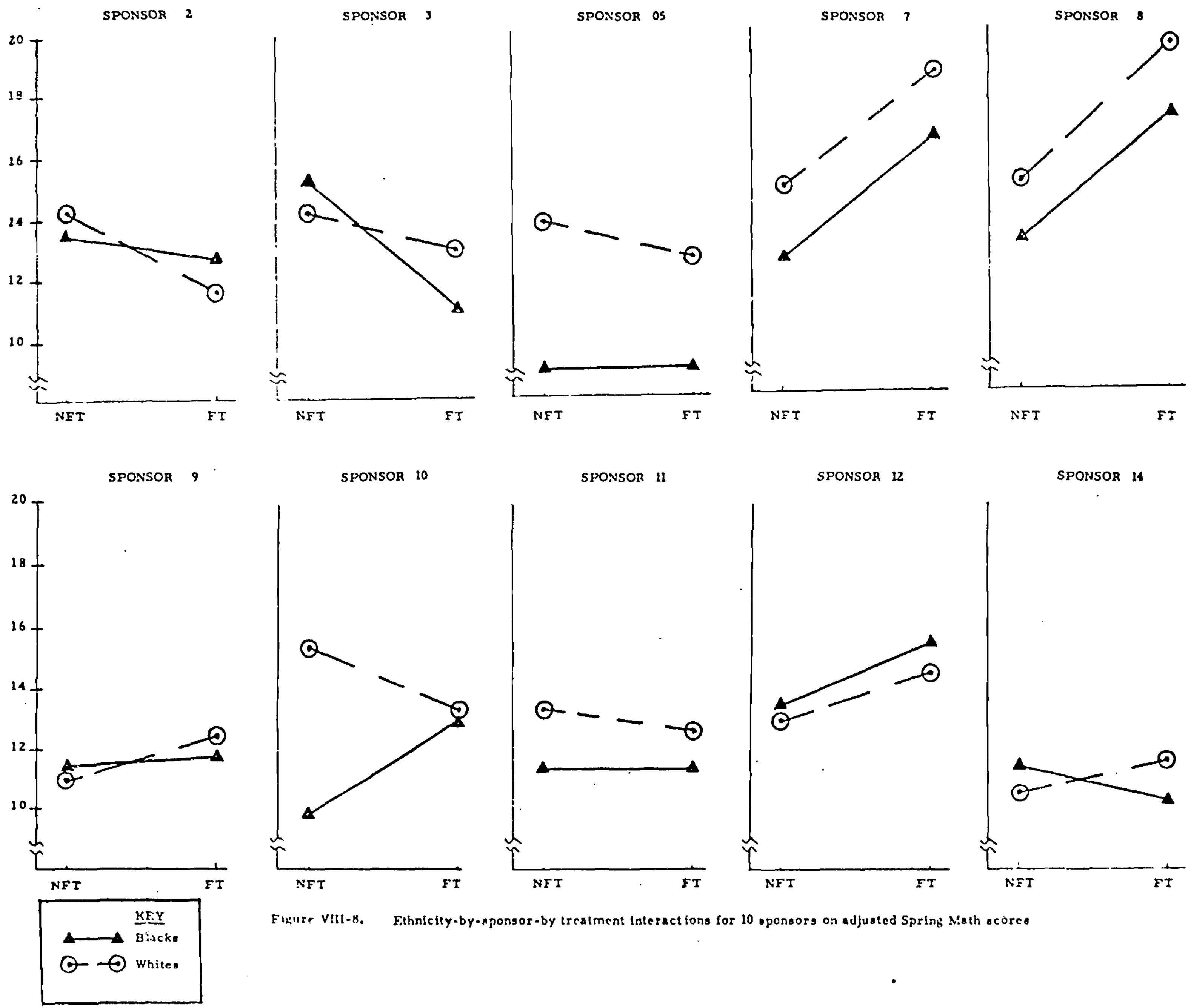


Figure VIII-8. Ethnicity-by-sponsor-by treatment interactions for 10 sponsors on adjusted Spring Math scores

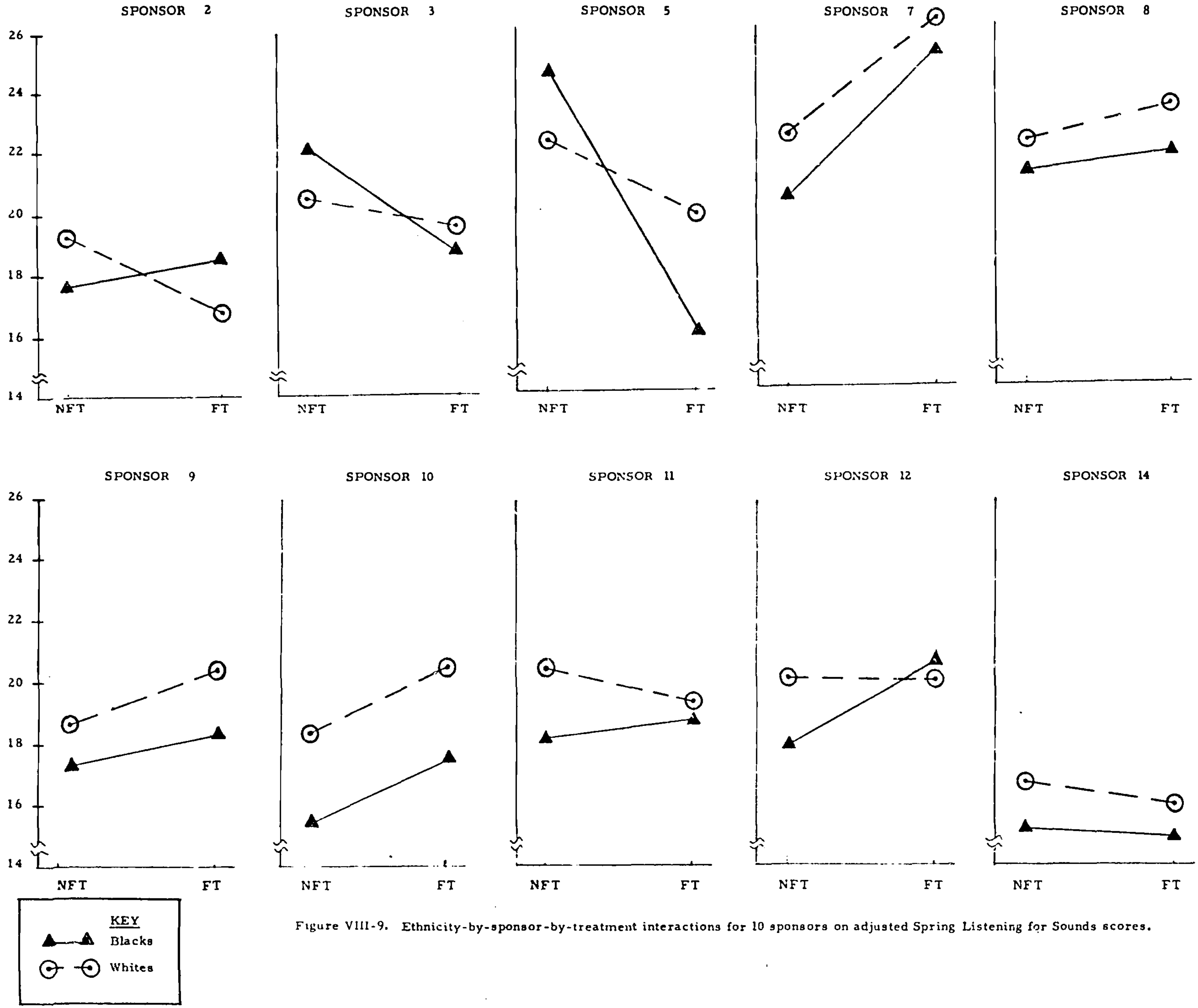


Figure VIII-9. Ethnicity-by-sponsor-by-treatment interactions for 10 sponsors on adjusted Spring Listening for Sounds scores.

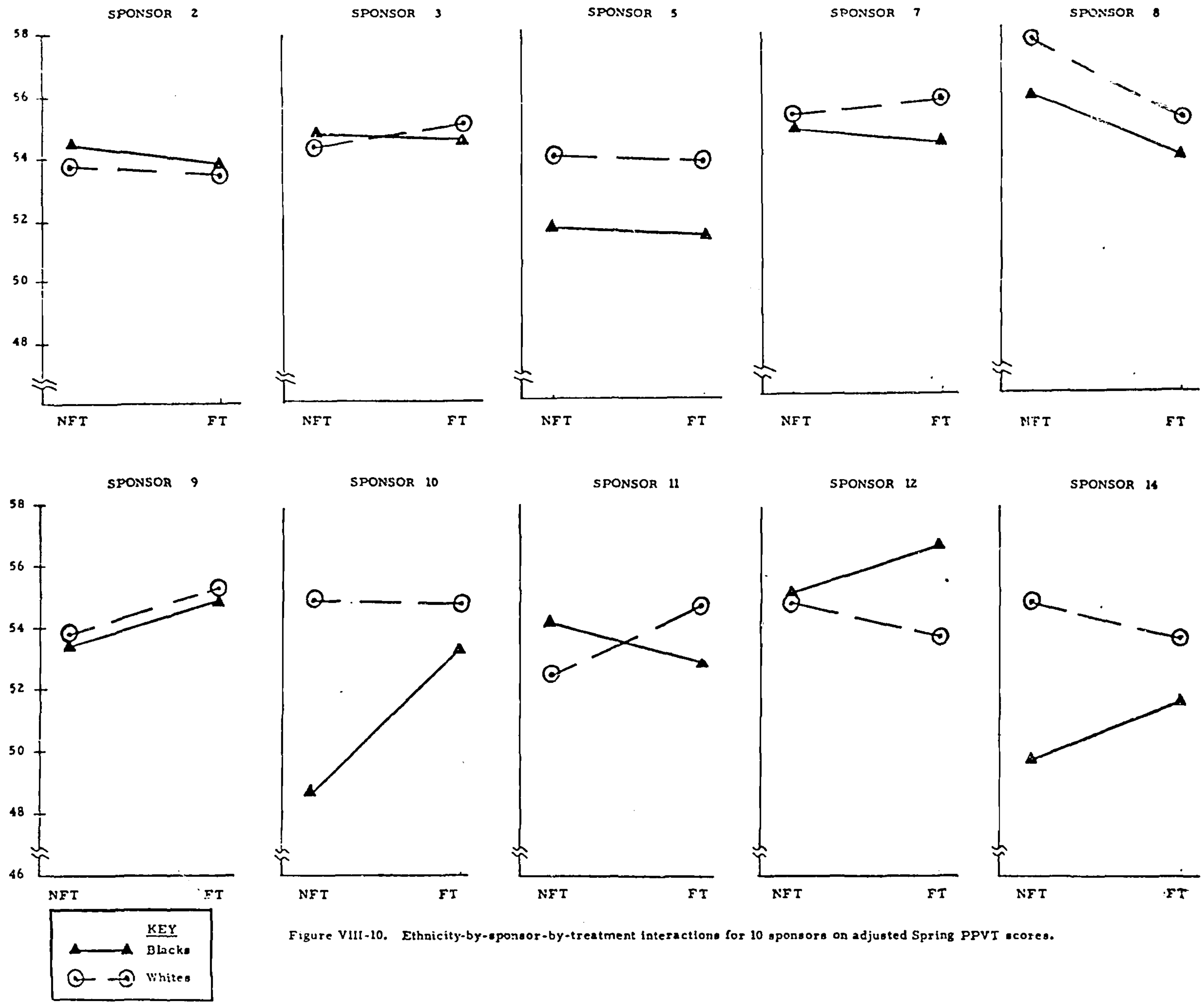


Figure VIII-10. Ethnicity-by-sponsor-by-treatment interactions for 10 sponsors on adjusted Spring PPVT scores.

clear that the Sponsors vary greatly in the patterns of FT/NFT contrasts between Blacks and Whites. For the University of Oregon (8), University of Florida (10) and University of Pittsburg (12), it appears that Blacks in Follow Through compare very favorably to NFT, while the Whites in Follow Through compare equally or slightly unfavorably to NFT. For Far West Laboratory (2) and Southwest Laboratory (14) there appear to be equivalent FT/NFT differences for both groups. For University of Arizona (3), Bank Street College (5) and Educational Development Center (11) it seems that FT Blacks are further below NFT Blacks than FT Whites are below NFT Whites.

If we consider next the PPVT interactions displayed in Figure VIII-10 we see that the patterns for the various Sponsors are not the same as for the WRAT. For the University of Florida (10), University of Pittsburgh (12) and Southwest Laboratory (14), the Blacks in FT show higher scores than NFT while the Whites in FT score equal to or slightly lower than their NFT groups. Far West Laboratory (2) shows the same PPVT pattern as the WRAT pattern of equivalent gains for all groups. Bank Street College (5), whose Blacks in FT compared more unfavorably to NFT than did the Whites on WRAT, demonstrates equivalent gains for all groups on the PPVT. On the WRAT, the University of Oregon (7) demonstrated very positive FT effects for both Blacks and Whites, but on the PPVT all FT and NFT groups showed approximately equivalent gains. For High/Scope (9) the comparisons on the WRAT and PPVT were just the opposite of University of Oregon's patterns.

Few Sponsors show similar patterns across all outcomes: the University of Pittsburgh (12) appears to produce higher FT/NFT differences for Blacks than for Whites on all measures except MAT-Arithmetic. The same is true for University of Florida (10) except for the MAT-Listening test where Whites and Blacks both show higher FT scores than NFT. University of Arizona (3) demonstrates less favorable comparisons of Blacks to NFT than of Whites to NFT on the four achievement measures. The University of Oregon (7) shows greater FT gains for both Blacks and Whites on three achievement measures, but on the PPVT, both groups show gains equivalent to the NFT group. In general, there appears to be considerable ethnic variation in FT/NFT within Sponsors across the

achievement areas, as well as across the Sponsors within any one achievement test.

4.4.0 DISCUSSION

From previous research we are led to expect ethnic differences in the school performance of children--differences which generally correspond with unfavorable social and economic forces acting on minority group members. Project Follow Through was designed, in part, to improve the school adjustment and performance of children from lower socio-economic levels. While there are many important economic factors, societal attitudes and home life-styles that are beyond the reach of most forces for educational change, Follow Through was intended to impact on parent, school and community structures as well as on the classroom context of the child. Yet, because of the diversity of approaches embodied in Follow Through and the variety of problems encountered, the search for the "best" model of educational change must at best be fatuous and at worst be dangerously misleading in a policy sense. A given program may be "best" only for certain kinds of children in certain types of situations. Ethnic differences are an important qualification of a program's effectiveness because they represent different types of forces acting on the lives of children, families and schools.

The results of this study suggest that ethnic differences may constitute a very real condition for differential Sponsor effectiveness. The variability in Sponsor effects between Blacks and Whites attests to this fact; however, the patterns that emerge across the Sponsors resist easy categorization by types of educational models. For example, Sponsors with highly structured curricula seem to produce equivalent positive FT effects in both Blacks and Whites on the MAT-Arithmetic test, but these Sponsors show different patterns on the WRAT and PPVT measures. On the WRAT outcome, the pattern of ethnic differences for a behaviorally structured model (University of Kansas - 8) resembles that of a model characterized by high parent involvement with relatively less emphasis on classroom programming (University of Florida - 10).

This study establishes ethnic variability in Sponsor effects, but for a number of reasons it cannot fully explain the conditions mediating

these effects. To date we have not related these results to specific differences in classroom processes across Sponsor models. Also, different motivational orientations in the children have not yet been considered in relation to academic performance. Equally important are the Sponsor differences in the types of parents and schools with which they work, and in the degree of rapport and commitment of the parents to the Sponsor's program. These factors affect not only the child's performance, but also the extent to which a Sponsor may implement a theoretical model. The present study does indicate, however, that Sponsor comparisons cannot accurately be made without regard to the conditions generating and pursuant to ethnic differences.

5.0 CHILD STUDY OF SEX DIFFERENCES

5.1.0 INTRODUCTION

Research on sex differences in academic performance and school adjustment has consistently revealed differences between boys and girls in motivational dynamics, affective orientations, psychomotor development, and academic performance. Most sex differences in school performance can probably best be explained in terms of differences in socialization patterns between boys and girls, although the tendency toward more rapid fine motor development and control in girls by the age of six years may also mediate superior female performance on reading and writing tasks (e.g., Pauley, 1951).

Sex differences arising out of socialization patterns are probably more germane, however, to the educational implications of Follow Through than are the psychomotor differences. Crandall and Rabson (1960) reported that while no sex differences in achievement motivation were measured in a sample of children of preschool age, boys in elementary school demonstrated higher achievement motivation than girls. Sex-typing in dependency behavior was described by Kagan and Moss (1962) in a longitudinal study which detected greater stability of dependency behavior in girls than boys between the ages of three and twelve years. As an example of the long-standing findings on sex differences in aggression, Jersild and Markey (1935) reported that four year old boys were more aggressive than four year old girls, but this was not true of two year olds. Also, boys were punished less for aggressive behavior than girls. The implication of these studies is that the acquisition of a number of important affective orientations basic to the child's style of adjustment is mediated by the sex role of the child.

In a longitudinal study of IQ growth and change, Sontag, Baker, and Nelson (1958) reported that the group of children whose IQ increased from the age of three years to twelve years included twice as many boys as girls. Increases in IQ were related to independence, competitiveness, and mastery achievement in school. In a survey of developmental studies, Bayley (1970) concluded that verbal scores stabilize earlier in girls than boys, but once the boy's verbal scores are established they remain more stable than for

girls. In terms of patterns of abilities, Bayley noted that mental abilities were more strongly intercorrelated in boys than in girls. Witkin (1962) reported a tendency toward more analytic reasoning styles in males than in females, and Anastasi (1958) refers to the superiority of males in numerical and spatial reasoning.

These sex differences in patterns of abilities are the result of complex interactions of socialization styles and nurturance patterns as well as genetic predispositions. In an attempt to explicate the relationship of some of these factors, Stanwyck and Felker (1971) analyzed the relationship of locus of control, self-concept, and anxiety in boys and girls at different elementary grade levels. Their results showed, in part, that girls tended to increase in anxiety from grades three to six while boys tended to decrease slightly. Girls tended also to internalize responsibility for success more than did boys. These relationships were not the same, however, for high self-concept and low self-concept groups. One explanation the authors suggested was that while the socialization and academic skill demands of the classroom favor girls at the outset, the girls tend to lose this advantage as they progress toward sixth grade and thus become more anxious. This was particularly true for girls with a low self-concept.

In summary, previous research provides ample reason to expect sex differences in the response of children to different educational models. Given the variety of classroom techniques in Follow Through--from open classroom to behaviorally structured to parent action programs--we may expect boys and girls to respond differently to them. Factors contributing to sex differences might be: the different nurturance styles of the teachers; variation in the range of independent activity, competition, and cooperation; differences in style of cognitive demands; and progression of the children from kindergarten to fourth grade. Because of theoretical model and program implementation differences, the Sponsors may be expected to vary on these dimensions. We cannot, however, describe with our current data the extent to which Sponsors differ on these variables from each other or communities and schools within any one Sponsor. Thus, although it is difficult to estimate the direction of sex differences for any one Sponsor,

the present analysis serves as an exploratory study to determine the extent to which sex differences may mediate Sponsor effects.

5.2.0 METHOD

5.2.1 Subjects

This study was conducted on the same sample of Cohort III kindergarten children included in the child level ethnicity study. Only subjects with complete Fall 1971 and Spring 1972 test batteries, as well as parent interview and teacher questionnaire data, were selected. Ethnic types other than Blacks or Whites were excluded because of the meag..r distribution across the Sponsors. The distribution of 3,830 subjects across the Sponsor, FT/NFT, and sex groups is indicated in Table VIII-10 together with group means for the entering WRAT scores, adjusted income, and proportion of mothers with at least a high school diploma.

5.2.2 Measures

The nine outcomes reported in the Head Start and ethnicity studies and discussed fully in Appendix A were also analyzed in this study. They include:

- Wide Range Achievement Test (WRAT)
- Peabody Pictures Vocabulary Test (PPVT)
- Metropolitan Achievement Tests
  - Reading
  - Numbers
  - Listening for Sounds
- Gumpgookies
- Locus of Control
  - Locus, positive
  - Locus, negative
- Absence

N = 3830			SPONSORS 2	3	5	7	8	9	10	11	12	14	GROUP MEANS
FT	Male	N = 1115	151	152	95	74	137	86	121	134	132	33	
		WRAT	36.17	34.01	31.91	36.15	33.64	34.06	31.71	34.01	36.29	28.82	34.09
		Adjusted Income	10.73	11.78	11.92	7.61	8.84	9.41	11.24	11.97	12.09	9.97	10.78
		Mother's Education	0.64	0.67	0.50	0.47	0.43	0.59	0.46	0.52	0.74	0.49	0.56
	Female	N = 1174	151	165	117	79	169	87	122	116	126	42	
		WRAT	38.64	37.75	37.01	39.38	38.50	34.07	35.15	35.06	40.98	34.07	37.41
		Adjusted Income	10.13	12.38	11.64	7.41	9.74	9.95	11.40	12.22	11.25	10.64	10.82
		Mother's Education	0.54	0.57	0.54	0.41	0.49	0.61	0.48	0.54	0.69	0.50	0.54
NFT	Male	N = 822	160	146	67	60	60	84	58	79	77	31	
		WRAT	33.65	37.69	37.54	33.42	27.77	30.95	32.57	38.49	35.96	27.65	34.34
		Adjusted Income	12.53	14.91	15.25	12.40	9.80	13.56	14.60	13.94	14.56	12.74	13.55
		Mother's Education	0.60	0.72	0.66	0.62	0.37	0.63	0.66	0.66	0.78	0.39	0.63
	Female	N = 719	143	119	64	59	50	60	43	68	83	30	
		WRAT	40.33	43.13	38.00	38.78	32.02	34.88	38.21	36.21	39.57	32.23	38.49
		Adjusted Income	11.64	16.47	15.14	13.78	10.66	12.72	14.63	13.88	15.39	10.63	13.73
		Mother's Education	0.55	0.77	0.66	0.51	0.46	0.48	0.47	0.56	0.77	0.47	0.60

TABLE VIII - 10

Distribution of subjects for sex by Sponsor by FT/NFT analysis with entering WRAT, Poverty Level, and Mother's Education group means

The covariates for this study were the same as those reported in the ethnicity study and fully discussed in Chapter IV. They include variables characterizing the child, parents and home, classroom, and community as follows:

- Fall WRAT
- Fall PPVT (used only for the PPVT outcome)
- Preschool experience
- Adjusted income
- Mother's education
- Years at current address
- School receptivity
- Parent-school involvement
- Teacher's years of education
- Teacher's years of teaching experience
- Percentage of White pupils in the classroom
- City size

5.2.3 Analytic Plan

The data from this study were analyzed with the multiple regression equivalent of a three-factor, fully crossed analysis of covariance. As outlined in Table VIII-10, there were ten levels of the Sponsor factor, two levels of the FT/NFT factor, and the two levels of sex. Details of this analytic technique are reported in the Methodology Appendix of this report.

The F ratios for the three-way interaction terms of sex by Sponsor by FT/NFT on each of the outcomes are computed with the following formula.

$$F = \frac{R^2_{Y\cdot cov,ABCDEFG} - R^2_{Y\cdot cov,ABCDEF}}{1 - R^2_{Y\cdot cov,ABCDEFG}} \cdot \frac{N - 50 - 1}{9}$$

The components of variance in the F ratio are defined as follows:

Analytic Model Components			df
cov (Covariates listed above)			11
A	Sex		1
B	Sponsors		9
C	FT/NFT		1
D	Sex	by Sponsors	9
E	Sex	by FT/NFT	1
F	Sponsors	by FT/NFT	9
G	Sex	by Sponsors by FT/NFT	9
			50

5.3.0 RESULTS

Since the purpose of this study was to determine differential sex effects across the various Sponsor by FT/NFT groups, the three-way interactions of sex by Sponsor by FT/NFT are of most interest. As reported in Table A VIII-3 of Appendix A, none of the F ratios for this interaction term were statistically significant on any of the nine outcomes. This result suggests that there were no differences attributable to the Sponsors' Follow Through effects in the manner in which boys and girls responded to the kindergarten year's instruction.

The main effect for sex assesses the extent to which there were overall differences in the kindergarten gains of boys and girls. Statistically significant sex main effects are noted in Table A VIII-3 of Appendix A for the WRAT, MAT Numbers, PPVT, Gumpgookies, and Locus, positive. Although statistically significant, the magnitude of the sex differences did not approach .25 standard deviation units of the outcome measure. The size of the overall sex differences indicated higher adjusted scores for boys by .12 S.D. units on MAT Numbers, .06 S.D. units on PPVT, and .04 S.D. units on Locus, positive; and higher adjusted scores for girls by .06 S.D. units on WRAT, and .07 S.D. units on Gumpgookies. It is generally felt that effects of less than .25 S.D. units do not demonstrate meaningful differences between groups. The

marginal differences noted above, however, on the WRAT and MAT Numbers scores represent a tendency consistent with previous studies demonstrating a slight female superiority in overall achievement and a slight male superiority in the Numbers area.

The descriptive data of group means in Table VIII-10 above, although not treated analytically here, indicate a general tendency in all but two Sponsor groups toward higher kindergarten entering achievement levels for girls. In this respect, the present sample is similar to those analyzed in other large studies (e.g., Bayley, 1970).

5.4.0 DISCUSSION

A brief review of previous studies on sex differences suggested that differences between boys and girls on motivational orientations, aggression, independence, mastery achievement, and other role expectations often lead to sex differences in academic performance and school adjustment. Considering the diversity of Sponsor approaches, ranging from behaviorally structured to open classroom to parent involvement programs, one expects sex differences to emerge in the performance and adjustment of children in these different educational environments. The data do not bear out this expectation. What factors may account for this apparent lack of effects?

The first consideration is the nature of the sample under study. If the characteristics of boys and girls in this sample do not correspond to those of previous studies, then expectations based on patterns from past samples may not generalize to this sample. The descriptive entry level data as well as the slight overall tendencies in kindergarten gains, however, suggest that the achievement differences between boys and girls found in past research apply also to this sample.

A second consideration is the nature of the Sponsors' models under study. Even though the Sponsors represent theoretically different educational models, it may be that the classroom dimensions most relevant to sex differences are common to a number of different models. For example, warm and nurturant teachers may be equally supporting to the child in a behaviorally structured and in an open classroom program. Also, it may

be that the general tendency toward greater permissiveness of aggression in boys is equally true in a number of different classroom environments, although the techniques employed for handling aggression differ. In a similar manner, sex roles for independence training may be pursued for boys and girls even though the techniques of expressing independence differ in various models. Although data describing Sponsors in such detail are not yet available, sex differences in Sponsor effects would not so likely be detected if these dimensions are distributed evenly across Sponsors in some form.

A third consideration is the extent to which the Sponsors are able to implement their programs in the various communities. It may be that a theoretical model could change patterns of sex differences, but that teachers incompletely trained or committed to the model maintain more traditional approaches to children. The implementation studies discussed elsewhere in this report suggest that the Sponsors experienced at times quite uneven success in establishing the concepts of their curricula in any one classroom or community. These kinds of variations, although not measured in the current sample, could attenuate potential sex differences expected from the general character of the Sponsors' models.

In summary, the lack of expected sex differences in the response of children to different Sponsors' programs leaves unanswered a number of issues which may obscure potential sex differences. Further studies in this area must include data describing in greater detail the quality of the teacher-pupil relationships, and must consider samples which clearly represent the Sponsor differences expected from their models.

6.0 CONCLUDING STATEMENTS

Although the major thrust of the evaluation of Follow Through is in the contrasts between the treatment and comparison groups, the search for the processes by which these effects are accomplished is of central interest to the educational and developmental specialists. The first step in this direction is to attempt to identify the conditions under which the FT programs are effective. These findings are not designed to be translated into policy decisions. It is not appropriate to decide that if certain Sponsors seem to be producing certain effects under specific

conditions that they should at this time be restricted to administering their models only under those conditions. This is a conclusion well beyond the data and one which can be drawn only after such findings are well replicated. However, if we determine the conditions under which some effects are found, we are in a position to ask why, and thereby develop an understanding of the dynamics underlying the phenomenon. This is our intent in these studies. The most impressive conclusion to be drawn from these findings is that there indeed appear to be specific conditions associated with many of the Sponsor effects. Because of the restricted samples available, many of these findings are very tentative, and will be followed up in future studies.

For example, it is of major interest to note that ethnically mixed classes show relatively greater scores on the motivational and absence measures (i.e., fewer absences in mixed classes). The social dynamics within these classes are not yet known, but they would be of very great interest to Sponsors. While we would be very interested in the effects each Sponsor has with mixed and non-mixed classes, only two Sponsors were sufficiently represented in the pool of mixed classes to allow this analysis. Both Sponsors showed significant effects with mixed classes. We do not know, however, if their effects are unique to these Sponsors or consistent across all Sponsors; we cannot draw any conclusions about the nature of the Sponsor impact.

At the same time, the distribution of high and low entry level classes across FT is rather different than that distribution across NFT. Sponsors may differ in the extent to which they are differentially effective with high and low achieving classes, but we cannot yet determine this from the current set of classes. It appears that some differences do occur, which would not be surprising given the potentially different approaches of the Sponsors. While we do not yet know why, there seems to be a slight trend in the direction of greater effectiveness with higher achieving classes among those Sponsors who show possible differences, a critical factor to consider when ultimately interpreting Sponsor effects.

The preschool study is a little more definitive because of a reasonable distribution of treatment conditions. However, the results

suggest some puzzles. The Sponsors who tend not to show higher FT achievement scores relative to their NFT groups generally (Far West, SEDL, Bank Street, and EDC) do begin to show such effects when they are involved with children who had some preschool experiences. On the other hand, while these Sponsors do tend to show some overall effects in the motivational measures, these effects are not found to be more pronounced in children with preschool experiences than in children without preschool experiences. These Sponsors, in other words, appear to show some achievement effects in the kind of children who have experienced preschool, while they tend to produce motivational effects in children with and without preschool. The University of Pittsburgh presents a very different pattern here. Whereas this Sponsor seems to be having very strong overall effects in the mathematics domain generally, children with Head Start experiences seem to be doing less well in mathematics than those without Head Start. It is necessary to know a good deal more about the kind of Head Start experiences which these children had in order to interpret this finding. In that manner, we may learn a great deal about the processes by which this Sponsor is producing the general methematic effects.

Much the same can be said for the findings in the ethnicity study. Black children participating in three different programs (Florida, Pittsburgh and SEDL) are uniquely responsive to the kind of instruction which leads to higher scores on the PPVT, but not to higher scores on the WRAT. Black children, however, respond to the High/Scope program with higher scores on both the PPVT and the WRAT. White children show somewhat different patterns of responses to these Sponsors, indicating that, to the extent that these groups of children have different instructional needs, techniques appear to be available which lead to the same kinds of achievement or motivational levels, albeit, perhaps, by slightly different routes.

In conclusion, there seems to be within the broad range of Follow Through programs, the kinds of resources which speak to the broad range of needs found within the many Follow Through groups. This implies

that the search for specific conditions under which each of the Sponsors may find their maximum effects is justifiable and potentially profitable. The next set of data will help establish the stability of these findings, and launch the search for the reasons why such phenomena are found.

CHAPTER IX

CROSS SPONSOR COMPARISONS: CONCLUDING REMARKS

1.0 INTRODUCTION

Although the preceding chapters should have led the reader to conclude that meaningful cross Sponsor comparisons are all but impossible, it is nevertheless the case that such comparisons will probably be of considerable interest to the educational community. The purpose of this chapter, therefore, is to deal more directly with these comparisons, as well as with the problems limiting the utilization of these first year findings for policy decisions. This chapter also presents recommendations for future research directions designed to move closer to the resolution of these problems.

At least two strategies might be employed in comparing Sponsor effects. The first is to classify Sponsors according to some specified set of dimensions (e.g., structured-unstructured, high-low parental involvement) and compare effects on each of the several measures. This strategy incorporates commendable features; however, it relies on data which permit a meaningful classification of Sponsors along the selected dimensions. The case has been made several times throughout this report that our knowledge of the reality of Sponsor operations is too meager at present to allow such classification. An alternative strategy is to sort Sponsors according to their patterns of effects with various kinds of children (e.g., preschool-no preschool, high entry level-low entry level). This strategy is consistent with the major goal of the Follow Through Evaluation: to determine what kinds of programs have what kinds of effects on what kinds of children at what points in time. Such an <u>ex post facto</u> approach, however, cannot be used to test hypotheses. Its primary role is to raise issues for future work, a role that is appropriate at this stage in the national evaluation.

In this chapter, then, we will group Sponsors by the effects found in the various studies carried out for this report. Since there are too many studies to synthesize into a single set of effects patterns, we have selected a few for summary purposes. The general plan is to examine Sponsor effects in terms of the following subject characteristics: (1) the ethnicity of the child, (2) the preschool experience of the child, and (3) the entry level of the class. In addition, selected results from the time of testing studies will be introduced in order to more fully describe Sponsor effects patterns.

The outcome variables chosen for inclusion in this chapter are the achievement test battery, including the MAT and WRAT, the PPVT, and the Gumpgookies test.

All data are drawn from the triple interaction studies, conducted at the child and class levels of analysis (see Chapter VIII), with the exception of the Gumpgookies effects. With respect to the Gumpgookies, since none of the triple interactions were significant, the data reported here are those derived from the analysis of school level main effects involving this instrument.

Let us turn now to a summary of the ethnicity of the child, by Sponsor, by FT/NFT findings.

2.0 ETHNICITY

Table IX-1 presents the Sponsors' adjusted effects on the Spring WRAT as a function of the ethnic membership of the children associated with each Sponsor. FT/NFT contrasts are summarized in the table to indicate whether the scores of the Sponsor's FT Black children were equal to, greater or less than those of the NFT Black children on the adjusted Spring WRAT. The same comparisons are made for each Sponsor's White children. Significant Sponsor x Ethnic membership x FT/NFT interactions were found for three other outcome measures: the MAT, Listening to sounds and arithmetic subtests and the PPVT which are also presented in Table IX-1.

First, we examine the pattern of Sponsor effects with Black and White children on the WRAT. Only University of Oregon shows positive effects (i.e., FT adjusted Spring WRAT scores are higher than NFT adjusted scores) for both Black and White children. On the other hand, EDC and Bank Street have relatively lower adjusted Spring WRAT scores for both Black and White children.

For Black children only, University of Kansas, University of Pittsburgh, and the parent education Sponsor (University of Florida) also appear to be producing high Spring WRAT scores compared to their NFT groups.

For White children only, the University of Arizona and High/ Scope also appear to be producing high Spring WRAT scores. On the

KEY TO THE SPONSORS

Sponsor 2: Far West Laboratory

Sponsor 3: University of Arizona

Sponsor 5: Bank Street College

Sponsor 7: University of Oregon

Sponsor 8: University of Kansas

Sponsor 9: High/Scope Educational Research Foundation

Sponsor 10: University of Florida

Sponsor 11: Educational Development Center

Sponsor 12: University of Pittsburgh

Sponsor 14: Southwest Educational Development Laboratory

Table IX-1

Sponsor Effects on Four Outcomes
Ethnicity Study

	WRAT			MAT--Listening to Sounds			MAT--Arithmetic			PPVT		
	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast
Whites	3,7,9	5,10,11	2,8,12, 14	7,8,9, 10	2,5	3,11,12, 14	7,8,9, 12,14	2,3,10	5,11	3,7,9, 11	8,12,14	2,5,10
Blacks	7,8,10, 12	5,11	2,3,9, 14	7,10, 12	3,5	2,8,9,11 14	7,8,10, 12	3,14	2,5,9, 11	9,10, 12,14	8,11	2,3,5,7

Table IX-2

Sponsor Effects on Three Outcomes
Head Start Study

	WRAT			MAT--Arithmetic			MAT--Reading		
	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast
Head Start	2,7,8,10	5	3,9,11, 12,14	7,8,9	3,12	2,5,10,11, 14	2,7,8,11, 14	5	3,9,10,12
Preschool	7,8,10	5	2,3,9,11 12,14	7,8,12	3,5,9, 14	2,10,11	2,7,10	5,14	3,8,9,11, 12
No Preschool	3,7,8,9, 12	5,11	2,10,14	7,8,12, 14	2,5,10	3,9,11	3,7,8,9, 10,12	5	2,11,14

other hand, the University of Florida shows relatively lower WRAT scores for White children, compared to its NFT group.

Table IX-1 also presents the findings for the MAT listening for sounds subtest. Once again, the University of Oregon shows positive effects for both Black and White children. This is also true for the University of Florida. On the other hand, Bank Street shows relatively lower scores on this measure for both Black and White children.

In addition to Oregon and Florida, the University of Pittsburgh shows relatively higher scores on the MAT listening subtest for FT Black children. On the other hand, the University of Arizona, like Bank Street, shows relatively lower scores on this subtest for FT Black children.

For White children, the University of Kansas and High/Scope join Oregon and Florida in having positive effects on the MAT listening subtest. Far West joins Bank Street in having relatively lower scores for FT White children.

Next, Table IX-1 presents the findings for the MAT arithmetic subtest. For both Black and White children, the Universities of Kansas, Oregon, and Pittsburgh all show higher scores on this subtest for their FT than for their NFT groups. For Black children, Florida also shows relatively higher scores on this arithmetic subtest, and for White children, SEDL and High/Scope appear to be producing higher scores.

Finally, Table IX-1 presents the findings for the PPVT. Here, the only Sponsor to produce higher scores for both Black and White children is High/Scope. On the other hand, Kansas is doing less well with both Black and White children on this instrument. For Black children, High/Scope is joined by Florida, Pittsburgh, and SEDL in producing higher PPVT scores. For White children, High/Scope is joined by Arizona, EDC, and Oregon in producing higher PPVT scores.

Combining these effects, several patterns seem to be emerging. The Universities of Florida and Pittsburgh programs seem to be having systematic positive effects with Black children in all of these outcome areas. Oregon is having systematic positive effects with White children

in the same outcome areas, and with Black children in all of these except the PPVT. Kansas has some effects on achievement with Black and some with Whites, but no effects on the PPVT with either group. High/Scope is effective with both Black and White children on the PPVT, and with White children on the other achievement measures. Several other Sponsors, covering a variety of approaches (i.e., Arizona, EDC, and SEDL) all have effects with some children in some areas. Bank Street and Far West appear not to have discernible effects in the achievement areas during the kindergarten year.

The diversity of these patterns needs to be emphasized. In the academic achievement areas, both the structured models and the parent education model are effective with Black children. On the other hand, with White children in the achievement areas, the cognitively oriented High/Scope and Arizona models, along with the Oregon model are effective. In the verbal, problem solving area (PPVT), High/Scope is very effective with both Black and White children. Finally, Kansas shows no effects and several other models show varying effects.

It is extremely difficult to generalize from these findings with the very limited set of information analyzed in this first annual report. However, with the limits of interpretation set forth elsewhere, it may be asserted that some of those models which aim directly at the kinds of skills measured on the WRAT (i.e., Oregon, Kansas, Pittsburgh) are in fact showing some real effects. Whether these effects are related to the particular kinds of children and communities with which these Sponsors become associated and whether they might be found for other groups of subjects, cannot be stated at this point. But the combination of structured classroom procedures with these particular target groups seems to be an effective set of events leading to relatively high scores on the WRAT.

At the same time, it is possible to assert that some of the Sponsors who have established a more indirect route to the skills measured by the WRAT (i.e., High/Scope and Arizona) are producing skills which are generalizing in a small but clearly discernible way to the performance of their children on the WRAT. Once again, the unique groups of children and communities with which these Sponsors are

involved, makes generalization beyond these data inappropriate. However, the combination of cognitive training in a responsive environment with these particular groups of children also appears to be an effective route to WRAT achievement.

Next it is also possible to assert that the parent education approach also appears to be producing positive effects on WRAT achievement. However, the dynamics at work in this approach are not clear, since little is known of the classroom events in these schools. Nor is a great deal known of the ways in which parent education processes are manifested in pressures on school personnel which may influence pupil performance. However, it is apparent that for some children, the parent education route to achievement on the WRAT leads to relatively high performance levels.

Finally, the Sponsors whose activities are farthest removed from those involved with WRAT-type skills (i.e., Bank Street and EDC) are showing little impact on WRAT scores in kindergarten. It should be obvious that such a lack of effects cannot yet be attributed exclusively to the nature of the model at this time. The particular properties of the children involved with these Sponsors and the very great difficulties faced in attempting major systemic change in school institutions (which is characteristic of these Sponsors) precludes any firm generalizations about model impacts.

Turning to the PPVT, a very different and more highly verbal instrument emphasizing receptive skills, the cognitively oriented Sponsor (High/Scope) shows consistent effects, and the direct achievement oriented Sponsors show diminished and variable effects compared to those produced with the WRAT. At the same time, some Sponsors from every category of model and program are showing some effects on the PPVT including EDC and the language development Sponsor (SEDL). Apparently there is also a variety of routes to improved performance on the skills measured by this instrument. The effectiveness of these routes also depends to some extent on the ethnicity of the children involved, so that the full meaning of Sponsor impacts cannot be determined until this issue is explored in the future.

3.0 HEAD START

Three outcome measures--Spring WRAT, MAT Arithmetic, and MAT Reading--produced significant Sponsor by FT/NFT by preschool experience interactions.

Table IX-2 (see page IX-4) presents the Sponsors' adjusted effects on the Spring WRAT as a function of the type of preschool experience--namely, Head Start (HS), other preschool experience (PS), or no preschool experience at all (NPS). The table shows whether each Sponsor's FT adjusted WRAT scores were greater or less than their NFT counterparts. Hence we can see which Sponsors are associated with high or low adjusted scores for each of these three groups of children.

Table IX-2 indicates that both the University of Oregon and the University of Kansas produce higher adjusted WRAT scores for FT children, regardless of whether or not they have had previous preschool experience of any kind. On the other hand, Bank Street has relatively lower WRAT scores for all three groups.

University of Florida produced higher WRAT scores with children with some form of preschool experience--be it Head Start or any other. Far West, on the other hand, has positive effects on the WRAT only with children with previous Head Start experience.

For children with no preschool experience, Arizona, High/Scope, and Pittsburgh join Oregon and Kansas in producing higher adjusted WRAT scores. On the other hand, EDC, like Bank Street, has relatively lower WRAT scores for children with no preschool experience at all.

Table IX-2 also displays the results of the three-way interactions for the MAT Arithmetic subtest.

Again the University of Oregon and the University of Kansas are the only two Sponsors associated with higher adjusted FT scores for all three types of children. Head Start graduates have higher adjusted scores in the FT programs of University of Oregon, University of Kansas, and High/Scope. The Universities of Arizona and Pittsburgh have lower adjusted arithmetic scores when we compare the FT Head Start graduates to the NFT Head Start graduates.

Children with a preschool experience other than that of Head Start

have higher adjusted arithmetic scores with Oregon, Kansas and Pittsburgh. The same type of children score low with Arizona, Bank Street, High/Scope, and SEDL. No FT/NFT differences on the MAT Arithmetic subtest were found for children with preschool experience for Far West, University of Florida, and EDC.

Children with no preschool experience obtain higher adjusted arithmetic scores with Oregon, Kansas and Pittsburgh, as well as the bilingual SEDL program. Far West, Bank Street and Florida have relatively lower adjusted arithmetic scores for NPS children.

Finally Table IX-2 presents the findings for the adjusted MAT Reading subtest scores. Here only University of Oregon produces higher adjusted reading scores for all classifications of preschool experience when we compare the FT to the NFT children. Bank Street is the only Sponsor with relatively lower scores for all children. As in MAT Arithmetic, the reading subtest has a pattern of Sponsor effects for the Headstart graduates which differs from that for the other preschool graduates.

Head Start children appear to be obtaining higher adjusted reading scores with Far West, Kansas, EDC, and SEDL, as well as Oregon. Bank Street has lower scores for the FT Head Start graduates than their NFT counterparts. Arizona, High/Scope, Florida, and Pittsburgh show no FT/NFT difference for these children on the adjusted reading scores.

Far West and Florida, like Oregon, produce higher adjusted reading scores for children with non-Head Start preschool experience. These FT children have relatively lower adjusted reading scores compared to their NFT groups with SEDL, EDC, and Pittsburgh, as well as Bank Street.

Children with no preschool experience are scoring well on reading with Arizona, High/Scope, Oregon, Kansas, and Pittsburgh, as well as Florida's Parent Education program. Only Bank Street has low adjusted reading scores with the FT children who stayed home before kindergarten.

Combining all patterns of Sponsor effects on the WRAT and MAT Arithmetic and Reading, we see that the University of Oregon and University of Kansas are the only Sponsors consistently having FT

adjusted scores higher than those of NFT. For Head Start children, Oregon and Kansas are effective on all three measures; for other preschool graduates only Oregon is associated with higher adjusted scores; for children with no preschool experience, Oregon, Kansas and Pittsburgh produce high adjusted scores.

There is a very clear trend in these findings which says that the highly structured, achievement-oriented Sponsors (i.e., Kansas and Oregon) are consistently effective with preschool graduates (as well as other children) and that the other kinds of Sponsors are having varying effects associated with the preschool experience of the child. While we know nothing of the kind of preschool experiences these children had, it appears to have prepared them for the kind of instruction they would receive upon entering the Kansas or Oregon Follow Through program. The preschool experience may have been one which prepared the children socially and emotionally for the kind of schooling they would face with these Sponsors (and not for the kind of experiences they would receive with the cognitively or developmentally oriented Sponsors). It also may have been a pre-kindergarten version of these achievement-oriented programs. If it was the latter, then we would want to examine the effectiveness of other Sponsors working with children who received experiences similar to those provided in Follow Through.

In other words, we would want to determine if it is consistency between preschool and Follow Through programs that produces these effects, or if it is something unique in the achievement-oriented programs which allows them to build upon the preschool experiences of these children. Still another possible explanation of this phenomenon is that the children who acquire the particular skills in preschool which allow them to respond to Follow Through are those who come from the kinds of families uniquely attracted to the achievement-oriented kind of program. Examination of the data collected in the Head Start Planned Variation study, when merged with the Follow Through data, will allow us to examine these alternative hypotheses. A clearer picture of this issue may be present when these analyses are reported in the next annual report of the national evaluation.

4.0 ENTRY LEVEL

Two outcome measures showed significant Sponsor x Entry Level x FT/NFT effects: Spring WRAT and MAT Reading. Table IX-3 presents the Sponsor's adjusted effects on the Spring WRAT as a function of the mean pretest achievement level of the classes associated with each Sponsor. FT/NFT contrasts are summarized in the table to indicate whether the Sponsor's FT low entry level classes were equal to, greater or less than the low entry level NFT classes on adjusted Spring WRAT. The same comparisons are made for each Sponsor's high entry level classes. Thus, it is possible to note which Sponsors were associated with relatively higher Spring WRAT scores when working with low entry level classes and which are associated with relatively higher Spring WRAT scores when working with higher entry level classes.

First, it is clear that only the University of Kansas produces higher Spring WRAT scores for both high and low entry level classes. That is, this Sponsor is relatively effective in producing WRAT scores regardless of the entry level of the class (as indicated by the parallel regression lines in Figure VIII-2). For the low entry level classes, a variety of Sponsors appear to be associated with higher adjusted WRAT scores than their comparison classes. These are: High/Scope, University of Florida and University of Pittsburgh. For the high entry level classes, Far West, University of Oregon join Kansas in showing higher adjusted WRAT scores in FT classes than in NFT classes.

Finally, Table IX-3 presents the findings for the MAT reading test. Here only the SEDL classes at both entry levels show higher MAT Reading scores regardless of entry level of the class. For the low entry level classes, those working with University of Kansas, High/Scope, University of Florida and EDC and also show higher adjusted MAT Reading scores. For the high entry level classes, Far West and University of Oregon also produce higher MAT Reading scores.

Combining these effects, it is apparent that University of Kansas, High/Scope and University of Florida are consistently effective with the low entry level classes, and that Far West Laboratory and

Table IX-3

Sponsor Effects on Two Outcomes
Entry Level Study

	WRAT			MAT--Reading		
	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast	FT Favoring Contrast	NFT Favoring Contrast	No FT/NFT Contrast
Low Entry Level	8,9,10 12	2,5,11	3,7,14	8,9,10, 11,14	2,5,12	3,7
High Entry Level[a]	2,7,8	11,14	3,5,9, 12	2,7,14	5,9,11	3,8,12

[a]Sponsor 10 is not shown here because of insufficient representation of high entry level NFT classes.

University of Oregon are consistently effective with high entry classes. Far West and High/Scope present very divergent patterns. The former shows consistently higher scores with the high entry classes and consistently lower scores with the low entry classes. High/Scope shows consistently higher scores with the low entry classes and generally lower scores with the high entry classes. Oregon and Kansas also show somewhat divergent patterns here. Oregon is consistently high with high entry classes and shows no effects with low entry classes whereas Kansas is consistently high for all classes. EDC and Florida are producing higher Reading scores in low entry classes, but Bank Street is consistently lower with these lower entering classes.

These patterns are very diverse and seem to indicate that there are complex factors associated with the entry level of the class. There is no doubt that teachers face different problems with and have different expectations of classes of varying entry levels. Such classes are also likely to differ in atmosphere and the expectations that children have of themselves. Further, it is also likely to be true that classes differing on achievement test scores at the beginning of the kindergarten year differ in a variety of other cognitive areas as well. The aptitudes with which the several Sponsor treatments are interacting are still unclear and unmeasured. It is necessary to examine the pattern of skills exhibited by the classes on entry level in order to know how to interpret the various patterns of Sponsor effects at the end of kindergarten. In addition, these entry level patterns need to be distributed across boys and girls as well as Black and White children in order to explore Sponsor effects fully. If there are enough cases to carry out these analyses in future data, they will constitute an important set of studies for the next report.

5.0 GUMPGOOKIES

In the preceding sections, we explored Sponsors' effects in the achievement domain. Sponsors were compared on the basis of the patterns of achievement effects they produced with various types of classes and children. The motivational domain is another important area of study, both for its own sake and as an important element in understanding the pattern of early effects which might be uniquely

associated with various Sponsors.

The less achievement oriented Sponsors might predict that their early effects should be in the affective domain, as a prerequisite to cognitive growth. On the other hand, the more achievement oriented Sponsors might predict that early achievement is a prerequisite to motivational growth, and should be apparent in those children for whom academic success enhances their sense of competence. It is conceivable, therefore, that a variety of effects on the Gumpgookies test of achievement motivation could be found among clusters of Sponsors.

There were no significant three-way interactions involving this instrument, indicating that Sponsor FT/NFT contrasts on this measure were not influenced by the kinds of categories into which each Sponsor's classes and children were classified for analysis. In order to examine Sponsor patterns on this motivational measure, therefore, it was necessary to consider the Sponsor FT/NFT main effects at the school level of analysis. Figures IX-1 and IX-2 summarize the Gumpgookies on the subset of schools excluding and including the Big Cities.

Contrary to expectation, there are no simple patterns of Sponsor effects on the Gumpgookies measure. All Sponsors except the University of Oregon and SEDL show higher Gumpgookies relative to their respective NFT schools (although EDC shows higher scores than their NFT schools only when the schools in the Big Cities are included). In general, these effects are rather large, ranging from .44 to 1.2 standard deviations higher scores for the FT groups than for the NFT groups. Oregon shows no effect on this measure, and the SEDL schools are about 1.0 standard deviations behind the NFT schools on the Gumpgookies.

Clearly there are multiple routes to higher scores on the motivational measure as well as some of the achievement measures. Sponsors who are producing higher achievement scores are associated with higher Gumpgookies as well as Sponsors who are showing no achievement effects at all. The fact that Oregon shows no effects on this measure, whereas Kansas shows significant positive effects, may

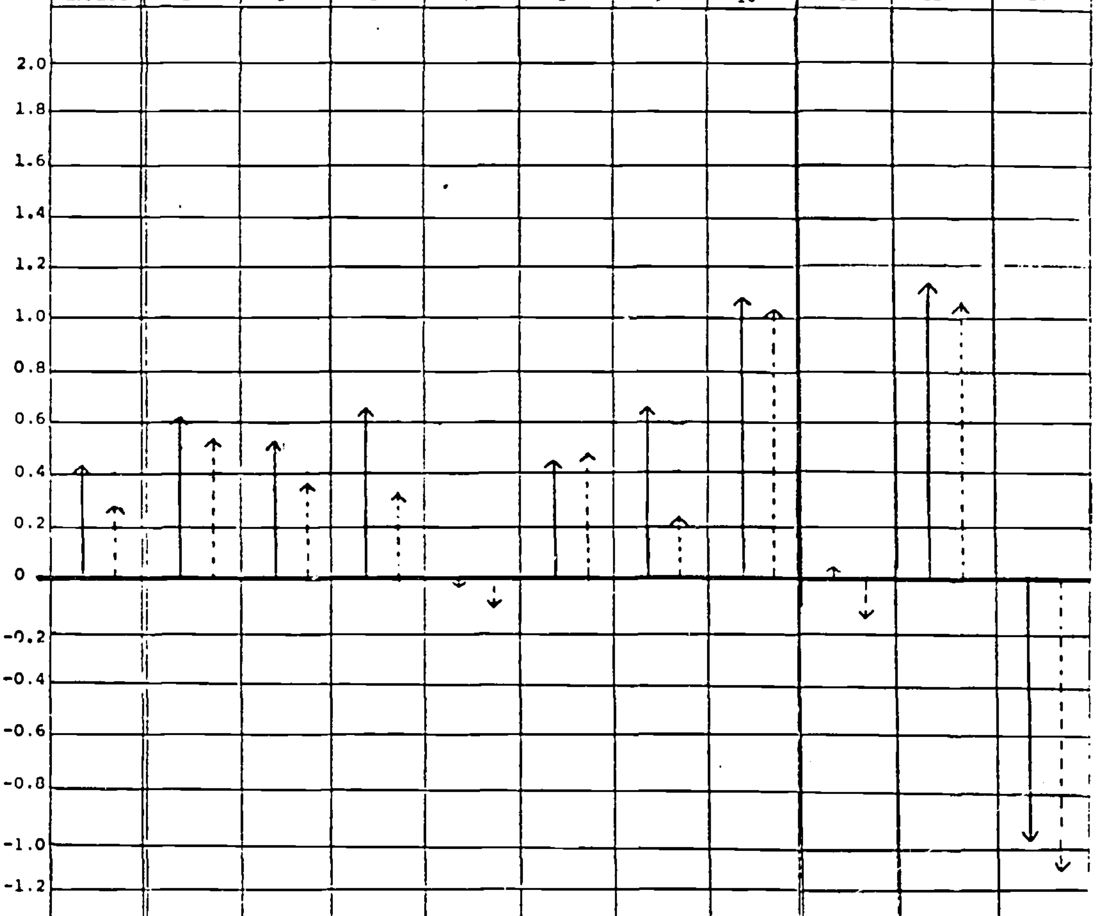
Figure IX-1

FOLLOW THROUGH EFFECTS PROFILE FOR GUMPGOOKIES

10-Sponsor School Population, Excluding Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)

KEY: Cov. adj. Effects (solid arrow); Unadj. Effects (dotted arrow)

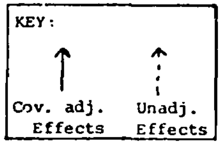


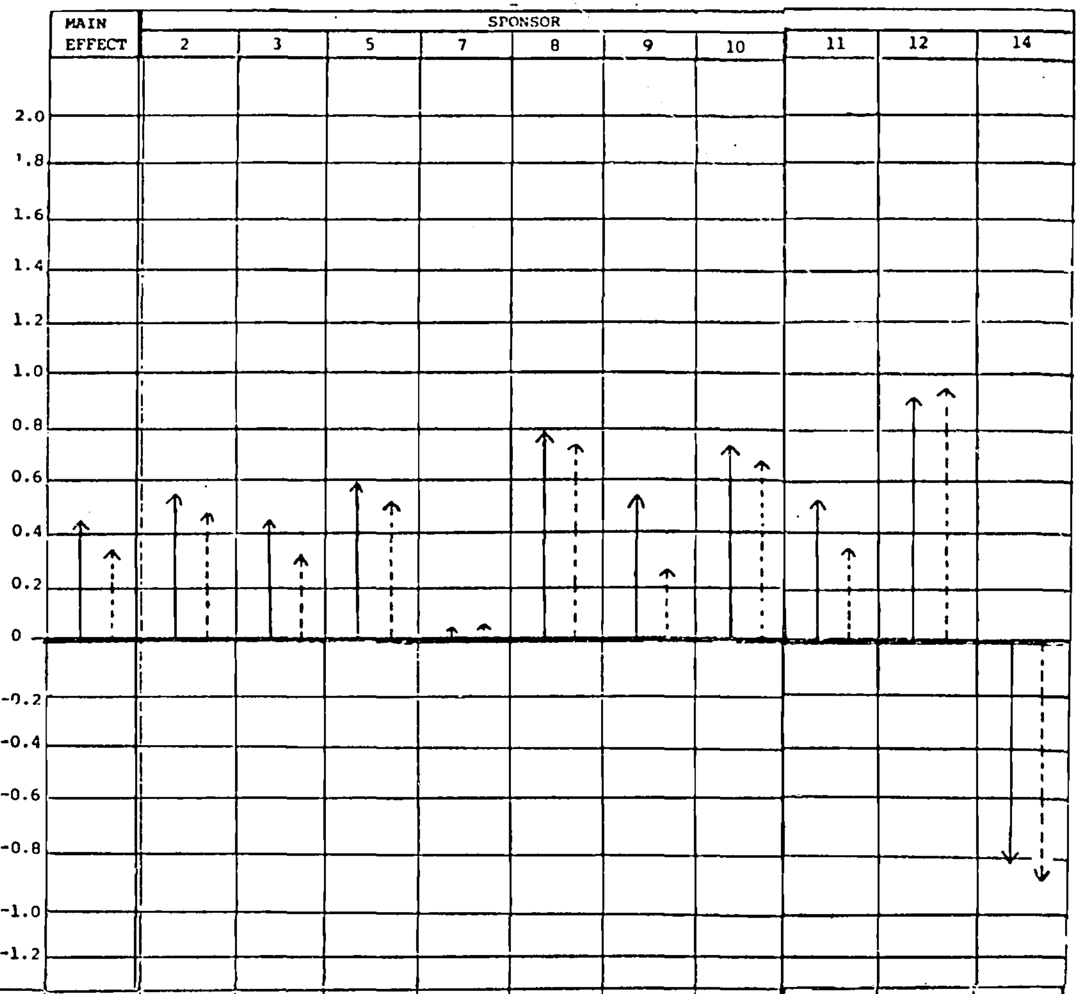
		MAIN EFFECT	SPONSOR 2	3	5	7	8	9	10	11	12	14
B = Magnitude of Effect	Adjusted	4.656	6.847	5.649	7.118	- 0.168	4.904	7.489	12.184	0.438	12.991	-10.886
	Unadj.	3.043	6.086	4.080	3.860	- 1.210	5.295	2.577	11.669	- 1.566	12.039	-12.398
SE = Standard Error of B	Adjusted	1.349	2.444	2.689	3.629	3.935	3.297	4.226	3.701	3.896	3.907	5.104
	Unadj.	1.303	2.743	2.949	4.041	4.242	3.604	4.586	3.980	4.337	4.242	5.715
t = B/SE = "Significance" Statistic	Adjusted	3.452	2.802	2.101	1.961	- 0.043	1.487	1.772	3.292	0.112	3.325	- 2.133
	Unadj.	2.335	2.218	1.384	0.955	- 0.285	1.469	0.562	2.932	- 0.361	2.838	- 2.169
S = Standard Deviation		11.279	11.279	11.279	11.279	11.279	11.279	11.279	11.279	11.279	11.279	11.279
*B/S = Effect in Standard Deviations	Adjusted	0.413	0.607	0.501	0.631	- 0.015	0.435	0.664	1.080	0.039	1.152	- 0.965
	Unadj.	0.270	0.540	0.362	0.342	- 0.107	0.469	0.228	1.035	- 0.139	1.067	- 1.099
N = Number of Schools in Computation	FT	137	29	21	12	11	16	8	15	10	9	6
	NFT	114	20	20	10	9	12	9	9	9	11	5

Figure IX-2
FOLLOW THROUGH EFFECTS
PROFILE FOR
GUMPGOOKIES

10-Sponsor School Population, Including Big Cities

*B/S: Magnitude of the Follow Through Effect in the Sponsor's Schools (in Standard Deviation Units)





		MAIN EFFECT	2	3	5	7	8	9	10	11	12	14
B = Magnitude of Effect	Adjusted	5.512	6.922	5.408	7.495	0.560	9.934	7.107	9.569	6.775	11.576	-10.222
	Unadj.	4.391	6.086	4.080	6.396	0.621	9.515	3.164	8.498	4.469	12.039	-10.955
SE = Standard Error of B	Adjusted	1.463	2.878	3.163	3.639	4.479	3.497	4.261	4.220	4.008	4.570	4.823
	Unadj.	1.456	3.387	3.641	4.188	5.091	3.980	4.864	4.803	4.665	5.237	5.662
t = B/SE = "Significance" Statistic	Adjusted	3.768	2.405	1.709	2.059	0.125	2.849	1.668	2.268	1.690	2.533	- 2.119
	Unadj.	3.016	1.797	1.121	1.527	0.122	2.391	0.651	1.769	0.958	2.299	- 1.935
S = Standard Deviation		12.530	12.530	12.530	12.530	12.530	12.530	12.530	12.530	12.530	12.530	12.530
*B/S = Effect in Standard Deviations	Adjusted	0.440	0.552	0.432	0.598	0.045	0.793	0.567	0.763	0.541	0.924	- 0.816
	Unadj.	0.350	0.486	0.326	0.510	0.050	0.759	0.253	0.678	0.357	0.961	- 0.874
N = Number of Schools in Computation	FT	156	29	21	16	11	20	11	17	13	9	9
	NFT	132	20	20	15	10	15	12	9	12	11	8

indicate that these two Sponsors are involving children in rather different ways in the activities of school. In this sense, they may not belong together in the same model category. On the other hand, the difference in effects may reflect differences in the characteristics of the families of the children associated with the two Sponsors or in the way in which they deal with these families. The fact that the parent education Sponsor (Florida) also has positive effects on the Gumpgookies suggests that this measure may in fact be influenced by factors external to the classroom and that, for purposes of clustering models, Kansas may have more in common with Florida than with Oregon along this dimension.

Among the Sponsors who show few effects in the achievement areas but strong effects on the Gumpgookies are the developmentally oriented Sponsors: Bank Street, Arizona, Far West, and, within the Big Cities, EDC. This suggests that the first step in the sequence which these Sponsors predict would lead to cognitive growth appears to be emerging; children in these schools seem to be willing to apply themselves to the school situation (as measured by the Gumpgookies). This may be a consequence of the opportunity to manipulate and explore their environment which these Sponsors intend to provide. It remains to be seen what the future course of growth is for these children, to the extent it can be measured by the present test battery.

The one Sponsor showing much lower scores on the Gumpgookies compared to their NFT schools is SEDL. It might be expected that the behaviors measured by this instrument are particularly diminished in the groups of Mexican-American children with whom this Sponsor works. However, the group of children included in this analysis were primarily Blacks and Whites whose cultural forms do not directly suggest that Gumpgookies scores might be low. Nor is there evidence that SEDL is producing lower Gumpgookies scores with just one group and not the other, since the triple interaction term involving ethnicity by Sponsor by FT/NFT is not significant. We are left with a quandary about this particular finding which will require much more intensive examination of the local site conditions to resolve.

6.0 TIME OF TESTING

As suggested in Chapter VII - 3.2, pupil test scores are influenced by the length of the instructional interval (the interval between pre and posttest administration across different Sponsors. Here we summarize Tables VII-9 and VII-10 to investigate the different patterns of effects found when each outcome measure is correlated with the length of the instructional interval, controlling for Fall WRAT scores. As explained in Chapter VII, the Fall WRAT correlated, in many instances, with both the pretest delay (the interval between the start of school and pretest administration) and the length of the instructional interval. The Fall WRAT was partialled out of these correlations to remove the effects of the non-random testing schedule. The patterns of effects which follow are not ordered in terms of subject characteristics but rather how the student scores are influenced by the length of the instructional interval.

The partial correlations of the Spring achievement measures with the length of the instructional interval, controlling for pretest differences, are positive and significant (at the .10 probability level) for four Sponsors: University of Arizona, Bank Street, EDC, and University of Florida. For these Sponsors, the longer the instructional interval, the higher the FT group scores on various achievement tests. The Spring WRAT is positively correlated with the length of the instructional interval for EDC; MAT Listening to Sounds for University of Arizona and EDC; and MAT Arithmetic for University of Arizona, Bank Street, University of Florida, and EDC.

Two of these Sponsors also have NFT groups with significant positive correlations between the MAT Arithmetic subtest and the length of the instructional interval: Bank Street and EDC. Although no Sponsor's FT group has a significant negative correlation between any of the achievement measures and the length of the instructional interval, Pittsburgh's NFT group has significant negative correlations for both the WRAT and MAT Reading measures.

The partial correlations of the Gumpgookies measure of achievement motivation with the length of the instructional interval are positive and significant for four Sponsors: Bank Street, SEDL, University of Kansas, and University of Florida. For these Sponsors, the longer the instructional interval, the higher the FT schools score on the Gumpgookies test.

TABLE VII - 9

Partial Correlations of Spring Scores
with Length of Instructional Interval
Controlling for Fall WRAT
FT Schools [a]

SPRING SCORES	SPONSOR 2 N=29	3 N=21	5 N=16	7 N=11	8 N=19	9 N= 9	10 N=16	11 N=13	12 N= 9	14 N= 9
Spring WRAT	.1323	.3163	.0088	-.2209	-.0204	.2582	-.0599	.6423*	.3079	.0579
MAT Listening	.0544	.6057*	.5692*	-.0032	.3148	.2852	-.1214	.6221*	.2285	-.5663
MAT Read	.2283	.4544*	.2010	-.5206	.0972	.0925	.2267	.5221*	.6158	-.2807
MAT Arith.	.3068	.5833*	.7101*	-.0994	.0649	-.0301	.4671*	.6600*	.2174	.1143
GUMP-COOKIES	.1497	-.7108*	.5073*	.3071	.6052*	.3788	.6040*	.4411	.0856	.7793*
LOCUS OF CONTROL Positive	.3432*	.1567	-.0031	.1377	-.0475	.4150	-.5531*	.4817	-.6971*	-.5531
LOCUS OF CONTROL Negative	.0759	.3306	.3124	.3586	.1839	.6656*	.0088	.1283	-.6746*	.0856

*These correlations are significant at the .10 level, using two-tailed significance tests.

[a] The sample of FT schools for this study was 152, four schools less than the total analytic subset. These four schools are not included here because of insufficient time of testing information.

Table VII - 10

Partial Correlations of Spring Scores
with Length of Instructional Interval
Controlling for Fall WRAT
NFT Schools[a]

SPRING SCORES	SPONSOR									
	2 N=20	3 N=20	5 N=14	7 N=10	8 N=15	9 N=10	10 N= 9	11 N=12	12 N=11	14 N= 8
Spring WRAT	.1153	.3198	.2631	.0953	.3128	-.4061	.4329	.1477	-.6257*	-.3440
MAT Listening	-.2257	.0496	.4394	.5301	.2649	-.1196	.1029	.4838	-.4007	.0502
MAT Read	-.0928	.1080	-.0364	.2351	.1904	-.0562	.0225	.0378	-.7641*	.3610
MAT Arith.	.1429	-.1583	.6774*	.2500	.3982	.1862	.3014	.6282*	.4723	.0114
GUMP-GOOKIES	.1825	-.1542	.4578	-.4549	.6929*	-.5783	.8709*	.6790*	-.3429	-.4971
LOCUS OF CONTROL Positive	-.0588	.2706	-.1933	-.1067	-.3239	.1280	-.4619	-.0835	.6388*	.2658
LOCUS OF CONTROL Negative	.1312	.2773	-.0187	.1581	-.3930	-.1214	-.5339	-.1461	.5446	-.4558

*These correlations are significant at the .10 level, using two-tailed significance tests.

a The sample of NFT schools for this study was 129, 3 schools less than the total analytic subset. These three schools are not included here because of insufficient time of testing information.

The Kansas and Florida Sponsors also have significant positive correlations for their NFT groups, as does EDC. Finally, Arizona's NFT schools have a negative partial correlation between the length of the instructional interval and Gumpgookies scores.

Positive correlations between the pre to posttest interval and Spring WRAT scores suggest either that there is an accumulation of achievement effects over the testing period or that achievement related events occur during the testing period which do not occur earlier. Since positive correlations are found for several Sponsors whose achievement effects are not very strong (i.e., Bank Street, EDC, Arizona, and SEDL), it may be that whatever effects occur for these Sponsors begin to emerge toward the end of the school year.

On the other hand, both Kansas and Oregon show no relationship between the length of the instructional interval and Spring WRAT scores (Oregon shows a negative but non-significant correlation), despite the fact that these Sponsors attempt to provide systematic sequences leading to accumulated success. Either the skills these Sponsors focus on do not generalize to the WRAT (which is not likely since their FT schools have much higher WRAT scores than their NFT schools), or the improvement in WRAT skills occurs at about the same time for all of these Sponsor's FT schools and remains on a plateau for the several weeks of the testing period. Another possible explanation for this finding is that a ceiling effect on the scores produced by these Sponsors may be present although this is not likely at the school level of analysis.

Although these correlations are based upon small samples of schools and need to be repeated on larger samples before stable conclusions can be drawn, the preliminary findings suggest that different Sponsors may produce different cognitive growth patterns at different points in time. Furthermore, the time sequence relating the acquisition of higher scores on the Gumpgookies test and the emergence of higher achievement scores for each Sponsor sheds additional light on these patterns.

For example, Bank Street, with generally low achievement scores, shows an increasing amount of achievement with instructional time, and also shows an increasing impact on the Gumpgookies with instructional time. This could imply a causal relationship between these variables, the consequences of which are beginning to emerge at the end of the kindergarten year. On the other hand, schools associated with EDC and Arizona show higher achievement scores, but no increase in Gumpgookie scores, with longer instructional time. This may mean that only those children with high Gumpgookies scores, which may have been observable by the middle of the kindergarten year, are beginning to respond to the EDC and Arizona models such that they can generalize their skills and attitudes to achievement test taking behavior. Finally, schools associated with Kansas are generally achieving higher than their NFT comparisons but show no increase in achievement with greater instructional time. At the same time, both Kansas' FT and NFT schools show increasing Gumpgookies scores with time. This suggests that for the communities associated with this Sponsor, the impact of school experiences on motivation may be a function of the types of children and the families from which they come rather than the nature of the school program. The higher achievement scores exhibited by the Kansas FT schools may suggest that this Sponsor has been successful in building upon this motivational property.

In sum, these data suggest very complex multivariate processes functioning within each Sponsor's group of schools. Processes such as these require considerably large sample sizes than those upon which these correlations are based. The possible relationships between Sponsor program, time of testing, and outcome domain will be examined in more detail as the data accumulate in sufficient quantity to justify appropriate analyses.

7.0 PLANS FOR THE FUTURE

The first step in the national evaluation of Follow Through has revealed some positive effects generally, and a multitude of patterns and suggestive findings. With the receipt of the next set of data (including the test scores for the first grade, Cohort III; second grade, Cohort II; and third grade, Cohort I), we are ready for the next series of analyses. These will focus on several new issues and the further examination of issues explored in this report.

The first of the new issues has to do with the interrelationships among the outcome variables. Multivariate techniques are available so that the patterns of achievement and motivational variables can be related to patterns of input variables. We wish to know how achievement and motivational variables relate to each other within each of the Sponsors; this can be examined with multivariate techniques.

Second, we wish to know the stability of such patterns over grades for the same Sponsors. Longitudinal tests of the stability of these patterns are now ready to be applied to the updated data base.

Next, it is necessary to deal with the problem of mismatch between FT and NFT for each Sponsor as well as between Sponsors. Several techniques for solving this problem are under consideration, including the generation of a "best matched" group based upon a careful search of the data base for an appropriate set of schools whose characteristics allow for reasonable contrasts.

The fourth issue has to do with the merging of parent, teacher, class, and school measures with pupil scores. This procedure will allow an opportunity to adjust more precisely among groups and to identify the contribution of these variables to pupil performance.

Next, it is imperative that data be collected on the implementation of Sponsors' models and programs at local sites. It is hoped that the preliminary efforts at describing these events can be expanded into a more systematic data collection process in a reasonable sample of sites. If this can be done, with quantification of these data, an estimate of the role of the programs can be generated. It will then be possible to

determine the contribution of the programs to pupil performance, and to parent and teacher measures. At the same time, we will be able to partial program effects out of the total pupil performance variance for an estimate of model effects.

Sixth, it is necessary to examine in greater detail with new data the issues surrounding relationships between time of pretesting, instructional length, time of posttesting and pupil performance. The data already suggest that both achievement and motivational effects become manifest at different times during the school year for different Sponsors, and these patterns need to be examined in much greater detail than has been done for this first report.

Next, the relationship between preschool experiences and Follow Through effects needs to be explored via the Head Start Planned Variation data. It will be possible with these data to determine the kinds of effects the several Head Start programs have produced and the persistence of these effects into kindergarten under several conditions. One major condition to be examined has to do with the consistency, in terms of Sponsor models, for children's Head Start/Follow Through experiences. The second condition has to do with the kind of children who have these experiences, and the ethnic mixes of the classes to which they are assigned when they enter Follow Through. Another condition of persistence of Head Start effects into Follow Through is in the kind of outcome domain through which the preschool effects express themselves. Many of these studies depend upon a sufficient group of children in the several analytic cells. As yet, we do not know how many children can be traced from Head Start to Follow Through, and this will determine the full range of HSPV studies possible. In any case several studies utilizing these data will be carried out.

All of the above issues speak to the problem of the FT/NFT contrasts under a variety of conditions utilizing a variety of measures. We are also interested in examining some theoretically important issues which go beyond the problem of FT effects. Thus, for example, we intend to explore some hypotheses having to do witn the kinds of children who benefit the

most from integrated classes. We are developing a theoretical model to generate hypotheses about the relationship between affective and cognitive development which are testable with the current data base. Hypotheses about the effects of low-achieving children in low- and high-achieving classes will be tested, as will hypotheses about teacher behavior in integrated classes with high and low academic performance levels. Some of these studies are now in progress, and they focus on the same data analyzed for this report. They will be replicated on the next set of data and reported in the next annual report. Other studies will be carried out when the full conceptual models are completed.

In any case, our activities in creating this first annual report of the Follow Through national evaluation have established a pattern for the next set of analyses and have generated a conceptual model and a set of hypotheses which are rooted in the questions raised by our findings to date. We shall report our hypotheses and the models underlying them as they are completed, but it is encouraging to report that the results of this first effort have led to new and more specific questions. Follow Through Planned Variation has produced more than some important findings. We also now know a little more about the important questions to ask, and this is what should be expected from a good experiment.

REFERENCES

Abt Associates Inc. Interim report. January 31, 1973, Contract No. OEC-O-72-5221, The United States Office of Education, Project Follow Through, 1973.

Anastasi, A. Differential psychology. (3rd ed.) New York: MacMillan, 1958.

Barber, T. X. & Silver, M. J. Pitfalls in data analysis and interpretation: A reply to Rosenthal. Psychological Bulletin (Monograph Supplement), 1968, 70 (No. 6, Pt. 2), 48-62.

Bayley, N. Development of mental abilities. In P. H. Mussen (Ed.), Carmichael's manual of child psychology, (3rd ed.), New York: Wiley, 1970.

Beller, E. K. Research on organized programs of early education, In R.M. Travers (Ed.), Second handbook of research on teaching, Chicago: Rand McNally, 1973.

Cohen, D. K. Policy for the public schools. Harvard Educational Review, 1968, 38, 114-137.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1968, 70, 426-443.

Coleman, J. S. et al. Equality of educational opportunity. Washington, D. C.: U.S. Government Printing Office, 1966.

Crandall, V. J. & Rabson, A. Children's repetition choices in an intellectual achievement situation following success and failure. Journal of Genetic Psychology, 1969, 161-168.

Dyer, H. S. School factors and equal educational opportunity. Harvard Educational Review, 1968, 38, 38-56.

Egbert, R. L. "Planned Variation of Follow Through", Paper presented at the Brookings Institution Panel on Social Experimentation, April, 1973

Flanders, N. A. Analyzing teaching behavior. Reading Mass.: Addison-Wesley, 1970.

Gray, S. & Klaus, R. A. The early training project: A seventh-year report. Child Development, 1970, 41, 909-924.

A guide to Follow Through. Washington, D. C.: U. S. Office of Education, 1973.

Hannan, M. T. Problems of aggregation and disaggregation in sociological research. (Methodology Working Paper, No. 4). Chapel Hill, N. C.: Institute for Research in the Social Sciences, 1970.

Hays, W. L. Statistics. New York: Holt, Rinehart & Winston, 1963.

Herriot, R. E. & St. John, N. H. Social class and the urban school. New York: Wiley, 1966.

Hoyt, C. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.

Hyman, I. A., & Kliman, D. S. First grade readiness of children who have had summer Head Start programs. The Training School Bulletin, 1967 63, 163-167.

Jersild, A. T., & Markey, F. V. Conflicts between preschool children. Child Development Monograph, 1935, No. 21.

Kagan, J. & Moss, H. A. Birth to maturity: A study in psychological development. New York: Wiley, 1962.

Karnes, M. B. Investigations of classroom and at-home interventions: Research and development program on preschool disadvantaged children. Final Report, Bureau No. 5-1181, Bureau of Research, Office of Education, U. S. Department of Health, Education, and Welfare, 1969.

Katz, F. I. Academic motivation. Harvard Educational Review, 1968, 38, 57-65.

Lord, F. M. Elementary models for measuring change. In C. H. Harris (Ed.), Problems in measuring change, Madison, Wis.: University of Wisconsin Press, 1967.

Lukas, J. & Wohleb, C. Implementation in Head Start planned variation 1970-71. Cambridge, Mass.: Huron Institute, 1972.

McPartland, J. S. The segregated student in desegregated schools. Baltimore: The Center for the Social Organization of Schools, 1968.

Monaghan, A. An exploratory study of the match between classroom practice and educational theory: Models in Head Start planned variation. Cambridge, Mass.: Huron Institute, 1973.

Pauley, F. R. Sex differences and legal school entrance age. Journal of Educational Research, 1951, 45, 1-9.

Porter, A. C. Analysis strategies for some common evaluation paradigms. Paper presented at the annual convention of the American Educational Research Association, New Orleans, 1973.

Porter, A. C. Some design and analysis concerns for quasi-experiments such as Follow Through. Paper presented at the meeting of the American Psychological Association, Honolulu, 1972.

Prescott, E. et al. Group day care as a child-rearing environment: An observational study of day care programs. Pasadena, Calif.: Pacific Oaks College, 1967.

Rosenthal, R. & Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston, 1968.

Rosenshine, B. & Furst, N. The use of direct observation to study teaching. In R. M. Travers (Ed.), Second handbook of research on teaching, Chicago: Rand McNally, 1973.

Rosenthal, R. & Jacobson, L. Pygmalion in the classroom: Teacher expectation and pupils' intellectual development. New York: Holt, Rinehart & Winston, 1968.

Scheffe, H. The analysis of variance. New York: Wiley, 1959.

Soar, R. S. Follow Through classroom process measurement and pupil growth (1970-71). Gainesville, Fla.: Institute for Development of Human Resources, University of Florida, 1972.

Sontag, L. W., Baker, C. T. & Nelson, V. L. Mental growth and personality: A longitudinal study. Monograph Soc. Res. Child Developm., 1958, 23, No. 68, 1-143.

Stanwyck, D. J. & Felker, D. W. Intellectual achievement, responsibility, and anxiety as functions of self-concept of third- to sixth-grade boys and girls. Paper presented at the annual meeting of the American Educational Research Association, New York, 1971.

Stein, A. Strategies for failure. Harvard Educational Review, 41, 158-204.

Weikart, D. P. Deloria, D. J., Lawson, S. A., & Wiegerink, R. Longitudinal results of the Ypsilanti Perry Preschool Project. Final Report. Ypsilanti, Mich.: High/Scope Educational Research Foundation, 1970.

Witkin, H. A., et al. Psychological differntiation. New York: Wiley, 1962.

Wolff, M. & Stein, A. Six months later, Head Start evaluation project. New York: Ferkauf Graduate School of Education, Yeshiva University, 1966.